UNIVERSITY
OF MICHIGAN
MAY 27 1955
SOCIAL SCIENCE
LIBRARY

Journal of Experimental Psychology

BF
1
J86

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

The Relation Between Abilities and Improvement with Practice in a Visual Discrimination Reaction Task: E. A. FLEISHMAN AND W. E. HEMPEL, JR. 301

Some Effects on Performance of Changes in Perceptual Load: R. CONRAD 313

Absolute Judgments as a Function of Stimulus Range and Number of Stimulus and Response Categories: C. W. ERIKSEN AND H. W. HAKE 323

The Effect of Familiarization upon Serial Verbal Learning: C. E. NOBLE 333

Effects of Three Types of Pretraining on Discrimination Learning in Preschool Children: G. N. CANTOR 339

Factors Influencing Rate and Extent of Learning in the Presence of Misinformative Feedback: R. E. MORIN 343

Transfer of Secondary Reinforcement Across the Hunger and Thirst Drives: M. R. D'AMATO 352

Effect of Monocular and Binocular Vision, Brightness, and Apparent Size on the Sensitivity to Apparent Movement in Depth: W. M. SMITH 357

Chronic Effects of a High-Frequency Stimulus on the Structure and Function of the Cochlea: I. E. ALEXANDER AND F. J. GITHLER 363

Effects of Training in Space Orientation on Perception of the Upright: M. WEINER 367

Taste Effects Resulting from Intermittent Electrical Stimulation of the Tongue: R. PIERREL 374

American Psychological Association

Vol. 49 No. 5 May 1955

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1955 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Department of Psychology, Stanford University, Stanford, California, until June 1, 1955. After June 1 address manuscripts to the editor at Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1955, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 49, NO. 5 MAY, 1955

THE RELATION BETWEEN ABILITIES AND IMPROVEMENT WITH PRACTICE IN A VISUAL DISCRIMINATION REACTION TASK

EDWIN A. FLEISHMAN AND WALTER E. HEMPEL, JR.

Skill Components Research Laboratory, Air Force Personnel and Training Research Center[1]

The present paper describes a further study concerned with the organization of abilities at different stages of practice in complex psychomotor learning. In a previous paper (3) the writers have shown that the particular combination of abilities contributing to individual difference in performance on a complex psychomotor task may change as practice on the task is continued. The issues involved in these studies bear on problems of predicting more advanced levels of proficiency in psychomotor skills as well as on questions concerning the processes involved in the learning of such skills.

In an earlier study, Reynolds (11) found a decrease in the correlations of certain printed test scores with performance on a psychomotor test (the SAM Complex Coordination Test, CM701E) as practice on this test was continued. He hypothesized that with training the pattern of abilities associated with the habits being acquired becomes increasingly specific to the task. The implication was that the best predictors of advanced psychomotor performance are intratask measures which reflect these specific abilities. In other terms, a large portion of the between-*S* variance at advanced levels of proficiency in such tasks may not be ascribable to variation along any other dimension than that established by the task itself.

Evidence somewhat contrary to this interpretation was subsequently presented by Adams (1) with the same Complex Coordination Test as the practiced task. Adams attributed Reynolds' finding to the particular class of variables selected to predict advanced psychomotor performance. Using psychomotor as well as printed tests, he was able to select a combination of independent measures which yielded a multiple correlation with advanced performance on the practice task that was even higher than the correlation between initial and final level of performance on the practice task itself. These results generally failed to support the hypothesis that

[1] This research was carried out at the Skill Components Research Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 7703. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

the abilities required for performance become increasingly specific as practice continues with a consequent decreasing predictability of performance. A number of tests maintained or increased their correlations from initial to final stage of practice. The problem appears to be a matter of identifying the abilities present and specifying the tests to measure them.

The present writers subsequently have attempted to specify more precisely the nature of the ability patterns contributing variance at different stages of practice on the Complex Coordination Test (3). More specifically, the techniques of factor analysis were applied to the intercorrelations among scores achieved at eight stages of practice on the practice task together with scores by these same *Ss* on 18 selected reference tests. The results indicated considerable, but systematic, changes in the factor structure of the practice task as practice on the task was continued. For example, certain "nonmotor" factors contributed most of the variance in earlier stages of practice, while certain "motor" factors contributed most of the variance at later stages. The rise in the acquisition curve was considered to be a resultant of systematic transformations in the particular combination of abilities used by *Ss* at different stages of practice. In other words, a given amount of practice may not result in an equal increase or decrease in favorableness for the operation of all the cooperating determining factors contributing to individual differences in performance on the task. This view is in agreement with that proposed earlier by Woodrow (12) in the case of practice on printed tests. The problem appears more crucial in the psychomotor area, however, since measures of such skills typically show quite rapid and extensive improvement with even very brief amount of practice.

The present paper describes a further study of this problem with a different psychomotor task and is aimed at testing the generality of certain of the previous findings. The practice task utilized earlier was a self-paced serial reaction kind of task. In the present study the task used was a paced visual discrimination reaction task, namely, the SAM Discrimination Reaction Time Test, CP611D2.

Method

Subjects

The *Ss* were 264 basic trainee airmen at Lackland Air Force Base. Each *S* received extended practice (see below) on the Discrimination Reaction Time apparatus, and in addition received a carefully selected battery of printed and apparatus reference tests.

Scheduling was arranged so that the printed reference tests were always administered to *Ss* between the psychomotor reference tests and the practice task. Half the *Ss* received the reference tests before the practice task while the other half received the practice task first. In the administration of the psychomotor tests, a rotational procedure was used in which the order of occurrence of each test in the series was fixed (in the order listed below), but different *Ss* started at different points in the series.

The Practice Task

The Discrimination Reaction Time Test, hereafter referred to as DRT, has been described fully by Melton (9). The test was designed to measure the speed with which individuals make differential manual responses to visual stimulus patterns differing from one another with respect to the spatial arrangement of their component parts. Briefly, the stimulus panel contains four lamps arranged in a square pattern. The upper left and lower right lamps are red and the upper right and lower left lamps are green. The test requires that *S* react by pushing one of four toggle switches (arranged on a horizontal base) as quickly as possible in response to the simultaneous lighting of a red and green signal lamp. The position of the lighted red lamp with respect to the lighted green lamp determines which of the four switches is correct. For example, if the lighted red lamp is to the right of the lighted green lamp, the right hand switch is pushed to

the right. If the lighted red lamp is below the lighted green lamp, the switch at the bottom of the response panel is pulled toward *S*. The four possible directions of movement involved in the task thus correspond in a certain sense to the four possible signal patterns formed by the lighting of the red and green lamps. The various stimulus combinations are presented successively in a fixed sequence of 20 settings, this sequence having been determined randomly.

If the correct switch is operated, a white lamp on the stimulus panel (which also lights up simultaneously with each red-green stimulus pattern) is extinguished immediately informing *S* of a correct response. The colored lights, however, do not go out until after they have been on for a total time of 3.0 sec., regardless of how quickly *S* has pushed the correct switch. The turning out of these lights at the end of each setting provides the ready signal for the next setting and marks the start of the foreperiod. The foreperiod before a new setting varies from .5 to 1.5 sec. in a randomly determined sequence. Between settings, *S* quickly returns his hand to a small cross at the front of the apparatus.

Before actual testing, each of the four signal patterns and associated corrected responses was demonstrated to *S* and this was followed by six practice signal settings. Efforts were made in the instructions for *S* to understand the basic principles involved. A test trial involved a series of 20 stimulus settings. Total score for a trial was the cumulated time (in .001 min.) taken to operate the correct switch for the 20 settings. The *S*s received 16 such test trials in a continuous session with a 40-sec. rest period between each 20-setting trial. Figure 1 presents the acquisition curve obtained for these 264 *S*s.

Brief descriptions of the printed and apparatus tests included in the study follow.[2] More complete descriptions of the printed tests included may be found elsewhere (5, 6). References to each apparatus test are given where available. The reliabilities of all these tests have been shown to be high in previous studies with comparable samples.

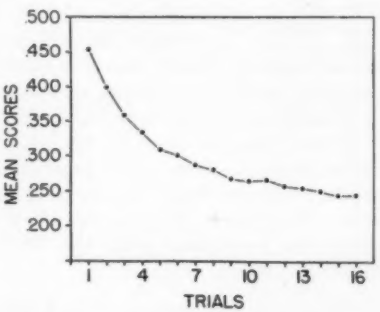


FIG. 1. Acquisition curve of performance on the Discrimination Reaction Time practice task.

Printed Test Variables

Word Knowledge, BI602A.—This is a multiple-choice vocabulary test.

Background for Current Affairs, BI102B.—This is an informational test covering current, recent, and historical events.

Mechanical Principles, CI903B.—Pictorial items require the comprehension of principles and mechanisms, such as leverage or rotation and transformation of motion, involved in the action and uses of various mechanical devices.

General Mechanics, BI902B.—Verbally presented items require practical mechanical information dealing with the use and operation of familiar mechanical methods and devices.

Tool Functions, BI904A.—For each tool presented pictorially, *S* indicates how it is properly used.

Speed of Identification, CP610A.—Pictorial items are presented in which the silhouette of an object must be identified when it is rotated and imbedded in a group of similar silhouettes.

Pattern Comprehension, BP803C.—A series of drawings which require visualization of relationships between components of solids and their unfolded flat projections is presented.

Instrument Comprehension, CI616C.—For each item which presents views of cockpit instruments, the examinee must determine the proper position or orientation of an airplane.

Visual Pursuit, BM802AX1.—From a series of mazes or irregularly curved lines, the task is to trace each line visually from its beginning to its proper termination point.

[2] Certain other tests also were administered, but were excluded from the present factor analysis on the basis of several considerations. This was done to reduce somewhat the computational labor without losing any basic information about the practice task. Thus, Numerical Operations, Log Book Accuracy, and Marking Accuracy (described in 6) and Rate Control and Single Dimension Pursuit (described in 9) were excluded because the correlations indicated extremely low communality with the practice task. Aviation Information, Electrical Information, and Dial and Table Reading (described in 5) were excluded because they appeared to duplicate, in impure form, factors already better represented in other reference tests.

Apparatus Test Variables

Purdue Pegboard—Assembly (10).—A series of small peg-washer-collar-washer assemblies must be completed as rapidly as possible. Score is the number of assembly components completed in two 60-sec. trials.

Santa Ana Dexterity, CM116A (9).—A series of square pegs with larger circular tops must be lifted, rotated 180°, and replaced in their respective holes as rapidly as possible. Score is the number of pegs rotated in four 35-sec. trials.

Rate of Movement (3).—The task is to break the beams between a series of photoelectric cells, one after another, by making scalloped movements of the hand as rapidly as possible. Score is the number of such movements completed in one 2-min. trial.

Rotary Aiming (2).—The task is to strike at a series of buttons arranged in a circular pattern on a horizontal panel going from one button to the next as rapidly as possible. Score is the number of strikes in four 30-sec. trials.

Visual Reaction Time (2).—The *S* keeps his finger on a button, merely depressing it ($\frac{1}{4}$ in.) in response to a single amber light before him. A click provides him with a ready signal before each light stimulus is presented with the foreperiod (between click and light) varying in a random order from .5 to 1.5 sec. Score is the cumulated reaction time for a series of 20 reactions.

Auditory Reaction Time (2).—Same procedure and scoring as Visual Reaction Time except that *S* responds to buzzer when it sounds instead of a light.

Jump Visual Reaction Time.—Same procedure and scoring as Visual Reaction Time except *S* does not keep his finger on the response button, but must move his hand 6 in. from a cross to the button as each light stimulus appears.

Jump Auditory Reaction Time.—Same as Jump Visual except response is to buzzer instead of light.

Complex Coordination, CM701E (9).—The *S* makes complex motor adjustment of stick and pedal controls in response to successively presented patterns of visual signals. Score is the number of completed matchings in one 8-min. test period.

Rotary Pursuit, CM803B (9).—The *S* attempts to keep a prod-stylus in contact with a small metallic target set in a rapidly revolving disc. Score is total time on target in 15 20-sec. trials.

Results

Data analysis.—For each test the obtained distributions of raw scores were transformed to normalized distributions of standard scores (stanines), each with a range from 1 to 9, a mean of 5, and an *SD* of 2. Conversions were made so that the 9 end of the scale was always indicative of "good" performance (e.g., low reaction time, high time on target, high number of corrects). Pearson product-moment correlations between the tests and practice task trials were then obtained.

Eight stages of practice on the DRT (representing every other 20-setting trial) were selected for inclusion in the factor analysis together with the 19 reference tests described above. The intercorrelations among the 27 variables are presented in the upper half of Table 1. Eleven factors were extracted from this matrix by the Thurstone Centroid Method (12). Factor extractions were continued beyond the point where any meaningful factor variance was suspected to be present. Table 2 presents the complete centroid factor matrix obtained. The bottom portion of Table 1 presents the final residual correlations remaining after factor extractions. It is apparent that only insignificant factor variance remains.

Orthogonal rotations of the axes defined by these factors were made using Zimmerman's Graphical Method (15). Table 3 presents the orthogonal solution of rotated factor loadings obtained using the criteria of simple structure and positive manifold. Factors were interpreted for psychological meaningfulness from the projections of the reference tests on the rotated axes. In general, loadings of .25 and above are consistent with the interpretations given.

Interpretation of factors.—The following interpretations are based on the loadings presented in Table 3. From the loadings of the first eight variables in

TABLE 1

INTERCORRELATIONS AND RESIDUALS*

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1. Trial 1, DRT		74	71	69	62	59	57	56	33	34	37	30	24	38	41	45	21	23	34	18	00	12	04	19	09	44	29
2. Trial 3, DRT	−01		82	78	74	68	66	64	36	36	40	39	27	25	44	42	23	28	47	24	11	14	00	26	24	48	28
3. Trial 5, DRT	−02	−01		83	79	72	72	71	38	40	37	37	29	24	44	47	25	32	41	28	16	17	12	24	22	45	30
4. Trial 7, DRT	−01	01	01		80	74	72	74	35	38	37	22	18	39	36	44	21	31	38	14	24	13	32	20	44	31
5. Trial 9, DRT	−01	−01	01	−01		77	73	77	30	36	35	24	20	37	37	42	20	34	36	30	19	26	14	35	30	41	31
6. Trial 11, DRT	00	01	01	−01	00		74	80	30	33	34	23	17	39	34	40	16	30	42	23	19	21	14	30	28	37	25
7. Trial 13, DRT	00	−02	00	−01	00	−01		79	31	37	39	24	19	43	40	38	17	34	42	23	15	20	14	31	27	43	27
8. Trial 15, DRT	−01	01	00	00	01	00	01		28	30	38	26	18	40	34	43	22	32	36	25	23	25	20	37	33	38	24
9. Word Knowledge	00	−03	00	−01	01	00	01	01		71	50	53	33	42	50	44	17	24	26	19	05	−04	02	11	09	23	15
10. Background for Current Affairs	01	−01	01	−01	01	00	02	01	01		50	48	34	48	55	45	16	24	25	18	03	−02	02	14	10	24	06
11. Mechanical Principles	01	01	00	01	−01	−01	−02	00	00	00		60	58	41	54	58	33	25	20	20	07	−09	−01	09	07	46	25
12. General Mechanics	01	00	01	01	01	01	−02	00	−02	00	00		71	27	41	47	23	19	19	16	06	−10	00	04	−04	29	21
13. Tool Functions	−03	−02	00	01	02	−01	02	01	03	00	−01	03		28	42	42	27	21	15	14	−01	−11	00	−03	−07	33	23
14. Speed of Identification	−02	01	01	01	01	02	−01	00	00	02	00	00	−01		52	45	32	46	38	16	25	08	06	23	25	39*	27
15. Pattern Comprehension	00	01	−01	02	−02	−01	01	−01	01	02	00	01	−01	00		51	31	36	27	19	07	−06	−04	08	08	40	23
16. Instrument Comprehension	02	02	01	00	−01	00	02	−01	01	02	−02	00	−02	01	00		32	33	29	20	13	04	05	19	14	47	24
17. Visual Pursuit	−01	00	02	−01	−01	01	02	−02	00	00	01	00	01	00	00	−03		27	20	17	23	03	−03	19	15	38	32
18. Purdue Pegboard Assembly	−02	01	−01	02	−01	01	01	01	00	02	−01	00	01	02	−01	−01	00		40	16	27	04	05	26	21	34	22
19. Santa Ana Dexterity	−01	01	−01	00	02	−02	01	01	00	01	00	01	02	−01	−02	−03	03	−03		24	31	14	08	30	30	38	35
20. Rate of Movement	−01	00	03	−01	−02	−02	01	00	−03	−01	−01	03	−03	03	−03	−01	00	00	02	03	32	21	17	34	30	24	24
21. Rotary Aiming	01	02	−03	01	02	01	−02	00	−02	00	−01	−04	01	02	−02	−01	−01	00	02	03	00	57	61	44	49	00	10
22. Visual Reaction Time	01	01	−02	00	00	−02	−01	00	−01	−03	01	−02	−02	−02	−01	−02	01	−02	−03	01	02	−02	00	44	49	00	15
23. Auditory Reaction Time	00	01	−01	00	01	−01	01	00	00	−01	02	00	00	−01	00	−02	01	−02	−01	−01	02	04	00	73	28	23	
24. Jump Visual Reaction Time	02	−01	00	−01	−01	−02	00	00	02	−01	00	01	−01	00	00	−01	01	01	00	00	−01	04	03	−02	73	19	20
25. Jump Auditory Reaction Time	04	01	−02	−01	−01	−03	00	−02	−02	−02	00	−01	−02	00	00	00	−01	01	01	00	01	04	03	−02		40	
26. Complex Coordination	00	−01	00	00	−01	00	−04	01	00	−02	00	−02	−02	00	00	02	−02	00	−02	02	−03	06	01	00	02		
27. Rotary Pursuit	00	−05	00	00	−02	01	00	−01	04	−02	00	00	00	01	01	02	−02	02	02	−02	−03	00	00	01	00	00	

* Decimal points have been omitted.
Values above the diagonal are the original correlations.
Values below the diagonal are the residual correlations after extraction of common factors.

TABLE 2

CENTROID FACTOR LOADINGS*

Variable	Factor											h^2
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	
1. Trial 1, DRT	67	29	20	−14	08	−24	−08	09	05	05	−12	69
2. Trial 3, DRT	77	23	25	−19	18	−15	−13	13	05	−06	10	85
3. Trial 5, DRT	79	26	28	−19	08	−12	−11	−09	−03	06	09	86
4. Trial 7, DRT	78	20	36	−17	04	−04	−08	−02	11	05	01	83
5. Trial 9, DRT	78	16	35	−16	06	09	−02	−05	04	08	06	81
6. Trial 11, DRT	73	20	37	−14	−10	15	10	−05	01	−05	−03	79
7. Trial 13, DRT	75	17	30	−15	−03	13	13	05	−16	−12	−11	79
8. Trial 15, DRT	76	13	35	−11	−17	19	18	−05	−07	−03	−01	84
9. Word Knowledge	54	27	−32	26	27	18	−13	−11	09	−10	−12	71
10. Background for Current Affairs	56	32	−26	28	33	16	−04	−15	06	−05	−07	73
11. Mechanical Principles	60	28	−38	15	−18	−07	16	07	14	−12	04	71
12. General Mechanics	49	32	−44	29	−24	07	−20	19	−08	10	04	79
13. Tool Functions	43	25	−46	19	−29	−12	−08	16	−18	08	18	69
14. Speed of Identification	62	−03	−23	−07	23	−03	16	−13	−12	−07	−03	56
15. Pattern Comprehension	59	21	−32	02	14	−11	14	−10	−04	−07	02	57
16. Instrument Comprehension	64	16	−25	07	−03	−11	13	−06	06	08	−02	54
17. Visual Pursuit	40	−14	−28	−14	−09	−16	11	−03	12	05	06	35
18. Purdue Pegboard Assembly	49	−16	−19	−18	14	09	15	−04	−22	15	06	46
19. Santa Ana Dexterity	54	−20	−07	−22	17	08	−13	19	−12	−08	−08	51
20. Rate of Movement	36	−14	−08	−08	08	19	−16	−02	20	−05	10	29
21. Rotary Aiming	32	−45	−11	−21	−12	23	−09	−21	03	−04	06	48
22. Visual Reaction Time	29	−48	41	31	−11	−12	−10	−12	−03	06	−20	67
23. Auditory Rection Time	24	−38	30	43	−10	−06	−11	−11	−17	−13	05	56
24. Jump Visual Reaction Time	50	−58	25	27	04	11	08	11	15	17	−03	80
25. Jump Auditory Reaction Time	44	−56	26	26	12	10	13	09	07	−09	15	72
26. Complex Coordination	61	−07	−17	−18	−10	−23	09	13	10	06	−13	56
27. Rotary Pursuit	44	−22	−12	−20	−16	−16	−13	10	03	−05	−07	38
$\Sigma a^2/k$	34	08	08	04	02	02	02	01	01	01	01	

* Decimal points have been omitted.

TABLE 3

ROTATED FACTOR LOADINGS *

Variable	Factor**											h^2
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	
1. Trial 1, DRT	*41*	11	04	22	*25*	00	10	15	*60*	−03	01	69
2. Trial 3, DRT	*47*	15	00	20	*28*	17	08	17	*58*	02	*26*	83
3. Trial 5, DRT	*56*	16	05	08	23	21	19	17	*56*	−14	13	86
4. Trial 7, DRT	*52*	22	00	12	21	*29*	09	07	*58*	−06	05	82
5. Trial 9, DRT	*52*	24	09	05	14	*39*	13	07	*52*	−04	01	80
6. Trial 11, DRT	*60*	23	10	11	13	*37*	15	01	*39*	08	−08	78
7. Trial 13, DRT	*62*	23	10	18	16	*30*	21	19	*31*	18	−07	81
8. Trial 15, DRT	*59*	*30*	15	10	07	*41*	23	04	*33*	10	−11	83
9. Word Knowledge	07	01	13	−02	*78*	15	13	10	15	04	−09	71
10. Background for Current Affairs	13	04	10	−11	*77*	11	21	08	17	07	−06	73
11. Mechanical Principles	02	02	*44*	*28*	*43*	11	*36*	−11	*27*	14	−02	70
12. General Mechanics	−09	00	*70*	04	*44*	09	08	18	*25*	04	−06	81
13. Tool Functions	−10	−01	*69*	09	*25*	−02	*26*	17	20	00	08	70
14. Speed of Identification	13	08	−05	13	*37*	18	*51*	23	16	04	00	55
15. Pattern Comprehension	13	−02	15	12	*46*	05	*48*	09	24	04	04	57
16. Instrument Comprehension	05	11	23	13	*36*	11	*41*	03	*37*	00	−08	54
17. Visual Pursuit	−15	02	07	*25*	07	18	*37*	02	*25*	−05	03	33
18. Purdue Pegboard Assembly	03	06	−01	−03	08	*27*	*46*	*35*	18	06	00	46
19. Santa Ana Dexterity	09	08	−07	24	17	*31*	10	*46*	20	12	07	48
20. Rate of Movement	04	06	06	13	16	*45*	00	02	13	−06	06	28
21. Rotary Aiming	−09	10	−07	15	−04	*57*	17	15	−07	−18	−02	46
22. Visual Reaction Time	−05	*72*	−13	13	−06	07	−07	10	01	−24	−20	68
23. Auditory Reaction Time	−08	*68*	05	06	04	05	−02	09	−20	−18	03	56
24. Jump Visual Reaction Time	−15	*73*	−16	05	03	*36*	05	11	21	10	−09	79
25. Jump Auditory Reaction Time	−05	*70*	−19	08	06	*33*	11	07	01	16	15	71
26. Complex Coordination	−01	09	08	*41*	10	14	*31*	15	*45*	04	−07	54
27. Rotary Pursuit	−04	09	07	*42*	01	21	12	22	23	−10	02	36
$\Sigma a^2/k$	09	09	05	03	09	07	06	03	11	01	01	

* Decimal points have been omitted.
** Factors are identified as follows: I—Discrimination Reaction Time Test Specific; II—Reaction Time; III—Mechanical Experience; IV—Psychomotor Coordination; V—Verbal Comprehension; VI—Rate of Arm Movement; VII—Perceptual Speed; VIII—Dexterity; IX—Spatial Relations; X—Residual; XI—Residual.

this table, the reader can trace the importance of each factor at different stages of practice on the practice task.

Factor I is common only to the stages of practice on the DRT and is not defined by any of the reference tests. For the present this factor will be designated the *Discrimination Reaction Time Specific*. It can be seen that intermediate and later stages of practice on the task have higher loadings on this factor than the early stages of practice.

Factor II is defined by the Simple, as well as the Jump Visual and Auditory Reaction Time Tests (Variables 22–25) and by none of the other reference tests. This confirms earlier findings (**2**) that *Reaction Time* is usefully considered a separate factor. It represents simply the speed with which *S* is able to respond to a stimulus when it appears. Table 3 also reveals that this factor is not involved in the DRT during the early stages of practice. However, there is a progressive increase in the importance of this factor as practice on the task continues until the final stage of practice where the loading is at a maximum.

Factor III is readily defined from its high loadings in the General Mechanics, Tool Functions, and Mechanical Principles Tests. It has been called Mechanical Experience in previous analyses (**6**). It does not appear to be involved at any of the stages of practice on the practice task.

Factor IV is identified chiefly by the Rotary Pursuit and Complex Coordination Tests. These tests have consistently defined a factor called *Psychomotor Coordination* in many previous analyses of the Aircrew Classification Batteries (**6**). The factor has been defined rather broadly as representing coordination of the larger muscles of the body in movements of moderate scope. Table 3 reveals that this factor is not involved in the Discrimination Reaction Time task at any of the stages of practice, except possibly to a negligible extent in the earlier trials.

Factor V is defined chiefly by the Word Knowledge and Background for Current Affairs Tests. These tests have been reference tests for a *Verbal Comprehension* factor in previous Air Force factor analyses (**6**). It appears from Table 3 that this factor makes a slight contribution to performance in the early stages of practice on the DRT, but that its importance decreases progressively as practice continues.

Factor VI is identified primarily from the loadings of the Rotary Aiming and Rate of Movement Tests. In previous analyses (**2, 3**) these tasks have identified a factor called *Rate of Arm Movement*, representing simply the speed with which such movements can be made. Its distinction from the Reaction Time factor has confirmed previous results (**2**).

As a side issue, the present study provided an interesting additional test of this. It will be recalled that two reaction time tests required only the pressing of a button (Variables 22 and 23), while two others (identical in other respects) required a hand movement of 6 in. to the button (Variables 24 and 25). As Table 3 reveals, the latter two tests appeared loaded on both the Reaction Time and Rate of Arm Movement factors, while the two simpler tests do not appear on the Rate of Arm Movement factor but only on the Reaction Time factor.

Table 3 indicates a rather dramatic increase in the contribution of this factor to the Discrimination Reaction Time Test as a function of practice. Loadings increase rather progressively from zero at Trial 1 to .41 at Trial 15.

Factor VII appears to represent a composite of at least two factors which failed to separate in the present analysis. For example, the Speed of Identification and Visual Pursuit Test are usually identified with a *Perceptual Speed* factor and Pattern Comprehension and Mechanical Principles have defined a *Visualization* factor (**6**). Whatever the nature of this factor in the present analysis, it does not seem very much involved in the DRT except perhaps for a very slight increase toward the latter stages of practice.

Factor VIII is defined from the loadings of the Purdue Pegboard and Santa Ana Dexterity Tests as a *Finger Dexterity* factor. This ability apparently does

not contribute to performance at any stage of practice on the DRT.

Factor IX appears to be a *Spatial Relations* factor. Most direct evidence of this is found in the high loadings of the Complex Coordination and Instrument Comprehension Tests as well as the Discrimination Reaction Time Test itself. These three tests have consistently been found saturated with the factor (3, 6). This factor seems to involve the ability to relate different responses to different stimuli, where either stimuli or responses are arranged in spatial order.

This factor is of special interest in connection with the DRT, since it has been found consistently to be the main factor measured by the test. As Table 3 reveals, the importance of this factor is maintained until approximately Trial 7. Then as practice continues there is a progressive decrease in its contribution to performance. In fact, in the later stages of practice other factors are contributing more variance than the Spatial Relations factor, which originally was the main factor in the test.

Ability patterns in relation to stages of practice.—These results generally confirm previous findings obtained with another complex psychomotor task (3). As before, considerable changes occurred in the particular combinations of abilities contributing to individual differences on the task as practice was continued. It was also evident that these changes were quite systematic and progressive during the learning period. The abilities contributing most of the variance later in training were not necessarily the same as those contributing at early stages of proficiency on the task.

Figure 2 illustrates this shift graphically. The figure presents the percentage of the total variance (factor loading squared) in performance accounted for by loadings on each factor at each of the eight stages of practice. This is indicated by the size of the shaded areas represented for each factor.

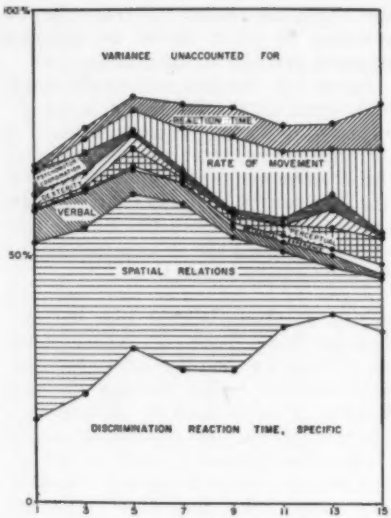


FIG. 2. Percentage of variance represented by loadings on each factor at different stages of practice on the Discrimination Reaction Time task. (Percentage of variance is represented by the size of the shaded areas for each factor.)

It can be seen in Fig. 2 that in the first stage of training, the abilities contributing more than 5% of the variance are Spatial Relations (36%), Verbal Comprehension (6%), Psychomotor Coordination (5%), and the Discrimination Reaction Time Specific (17%). In the last stage of practice, the abilities contributing are Rate of Arm Movement (17%), Reaction Time (9%), Spatial Relations (11%), Perceptual Speed (or Visualization) (5%), and the Discrimination Reaction Time Specific (35%).

Thus, early in practice the main ability contributing to individual differences on the task was Spatial Relations. By the last stage of practice, however, the variance contributed by this factor decreased from 36% to 11%. On the other hand, other factors, not present early in practice, were now contributing a combined total of over 30% of the variance in

performance, with principal contributions by Reaction Time and Rate of Arm Movement. The contribution of the factor common only to the task itself had increased from 17% to 35%.

The relative importance of different abilities at various stages of proficiency on the DRT may be represented in more traditional terms. For example, the Word Knowledge Test was considered to represent a measure of *S*'s Verbal Comprehension. The 264 *Ss* in the present study were then stratified into quartiles on the basis of

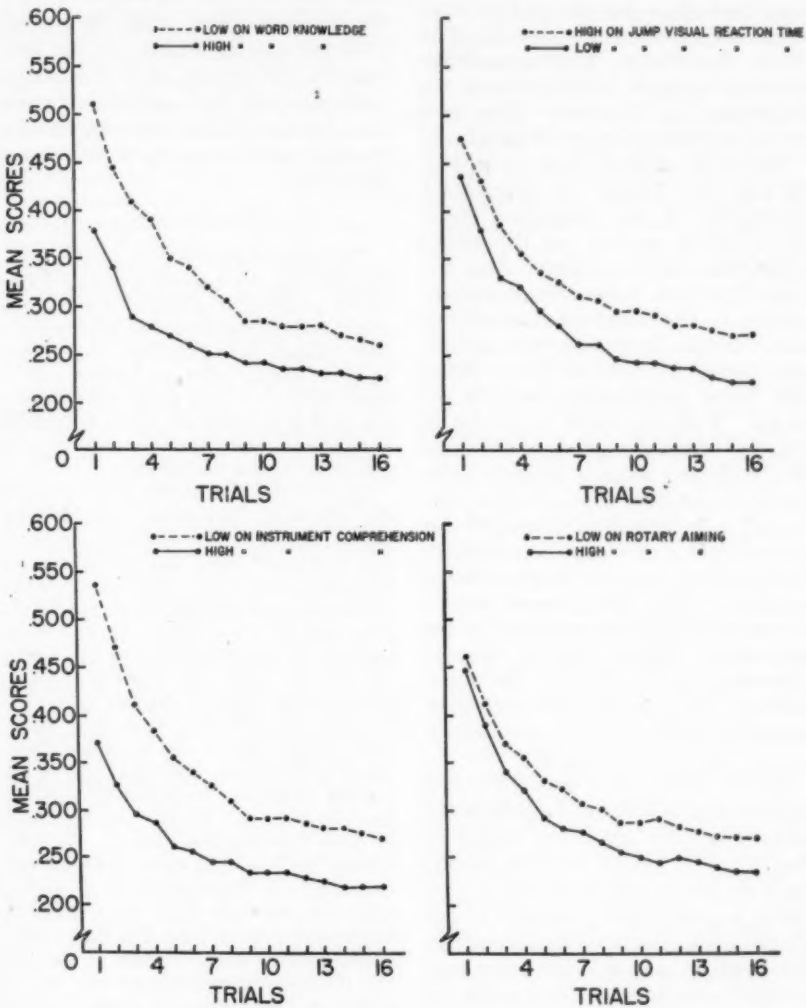


FIG. 3. Comparison of Discrimination Reaction Time acquisition curves for groups stratified on different predictor variables.

their scores on this independent measure. The mean scores achieved by each of these groups at each trial of practice on the DRT were then obtained. Figure 3*A* compares the acquisition curves for this task for the *Ss* in the upper and lower Verbal Comprehension quartiles. It can be seen that these curves tend to converge as practice is continued. The Word Knowledge Test apparently provides greater differentiation in Discrimination Reaction Time performance earlier in practice than later.

This is in marked contrast to Fig. 3*B* and 3*D*. Figure 3*B* represents a comparison of the acquisition curve for a group scoring in the upper quartile on the Jump Visual Reaction Time Test with the curve for a group scoring in the lower quartile on this test. In this comparison the acquisition curves for Discrimination Reaction Time performance tend to diverge. A similar result is evident in Fig. 3*D* where the group is stratified on the Rotary Aiming measure (representing the Rate of Arm Movement factor). In both Fig. 3*B* and 3*D* our independent ability measures differentiate later training better than earlier training. Figure 3*C* presents similar curves for the group stratified on the Instrument Comprehension measure (representing the Spatial Relations factor). Here we see greater differentiation early, but still considerable separation later in training.

These results, of course, are just another way of representing the factor analysis results presented earlier in Fig. 1. The factor analysis results have an advantage, however, in providing a more "pure" estimate of the contribution of each ability dimension. For example, any one test score may actually represent a composite of several factors. The factor results allow a partitioning of this variance into relatively independent factors and a separate evaluation of the importance of each factor in the practice task. Stratifying on a single test score (which may be factorially complex) may obscure some of the findings. Of course, there are still some difficulties in the application of factor analysis to such problems especially with regard to the resolution of very low loadings which might be shifted from one axis to another in the final rotation. However, as in any statistical approach, there are residual areas of uncertainty which are resolved only through replications in similar situations.

Discussion

The present results give further indications that the quantitative and qualitative pattern of abilities determining proficiency in psychomotor performance may change as a function of practice. In other words, individual differences in performance on such tasks after certain amounts of practice are likely to depend more on certain abilities and less on others than they did initially.

Some of the data of this study support the specificity hypothesis provided by Reynolds (11). For example, as practice continued there was an increase in the contribution of a factor found specific to the task and not defined by any of the other reference variables. However, the increase in the importance of this factor was not as marked as was found previously in the case of another task—the Complex Coordination Test (3). Moreover, in the case of the Complex Coordination Test there was a decrease in the *number* of factors contributing variance later in practice. In the present study no such decrease was evident. In fact, during the first trial three different factors achieved loadings of .25 or above, while in the last trial, four different factors achieved at least this magnitude. These results would appear to present still further evidence against any simple specificity hypothesis. These results tend to support the view by Adams (1) that the problem is more one of identifying the particular abilities important at later stages of practice.

A related problem is the need to determine the nature of the factor found specific to the stages of practice on such tasks. Such a specific factor was found in the Complex Coordination Test as well as in the present Discrimination Reaction Time Test. Moreover, in each test the "specific" was most strongly involved at later stages of practice. Is the "specific" factor found in each of these separate studies really specific to each task, or does it represent, at least in part, a more general factor common to advanced stages of practice on several tasks? Adams (1), for example, found that measures from advanced stages of training on several complex psychomotor tasks correlated more highly with each other than measures from different stages. Moreover, the best predictors of advanced performance were measures of advanced performance on other psychomotor tasks. This would seem to indicate the presence of common factors at advanced stages of practice on several tasks not held in common between these tasks earlier in practice. It is possible, for example, that this factor represents a kind of "integration" ability, which does not contribute variance until the component activities are well learned. Adams has hypothesized this to be a "patterning" response which involves the organization of the component responses into the proper sequence and pattern. It may be that individual differences in this ability are relatively independent of the abilities associated with learning the component responses.

Thus far it has not been possible to subject these hypotheses to crucial tests. Future analyses should be designed to extract possible common variance found at advanced proficiency levels and to define it. It is felt that the factor found specific only to the trials of the practice task may be broken down further with additional experimentation including other reference variables. A possible lead along these lines is furnished by a recent factor analysis study by the authors (4). In this study a factor called "Integration" was identified in a large battery of psychomotor and printed tests, quite different from the present battery. This factor was common to printed as well as apparatus tests in which *S* was required to utilize and apply information provided by a number of separate symbols, cues, or responses, into an integrated single response. It would appear desirable to include the best tests of this factor in subsequent factor analyses of advanced levels of psychomotor proficiency.

The possibility also exists that this "specific" factor represents "learning" ability of some kind. However, there was no way to test this in the present study. Woodrow (14) has investigated this hypothesis in the case of extended practice on printed tests. He included "gains" scores on a number of tasks as variables in the factor analysis. His analyses failed to reveal any general "learning" ability common to the various tasks. Instead, he found that loadings of practice gains scores are often specific to the particular task.

There is nothing in the results of the present study to alter the view presented previously by the writers (3) and still earlier in similar form by Woodrow (13). Performance at any stage of practice is regarded as determined by a set of cooperating but independently variable abilities. Continuation of practice on the task is regarded as a change in the conditions under which these various abilities operate. This change in conditions may involve several things. For example, in the early stages of practice on the DRT, learning the spatial characteristics of the stimuli and the responses, and the relationships among them, may be the chief contributor to individual differences. Understanding the problem posed by the instructions (Verbal Comprehension) also contributes to some degree. Once these are learned to a sufficiently high degree and no further improvement in these skills is possible (or necessary), within the requirements and limits of the task, other features of the task may assume increasing importance (e.g., reacting quickly and moving the hand as rapidly as possible). However, the important point is that earlier in the learning period the use of certain skills is a minimal but necessary requirement for achieving a certain amount of progress

in performing the task. At this stage, individual differences in these skills affect the rise in the acquisition curve. Later in training, individual differences in other combinations of skills may be the chief contributors to performance. In a sense, from the point of view of *S*, the nature of the task changes. Some may prefer to conceive of these changes in conditions as possibly involving differences in the rate in which inhibition or habit interference develops with regard to certain responses, or the rate at which certain responses become extinguished or reinforced as practice is continued. (For a recent series of studies on individual differences in susceptibility to interference and related factors see **7, 8.**) In any sense, the rise of the performance curve may be considered to be a resultant of systematic transformations in the particular combinations of abilities used by *S*s at different stages of practice. In other words, a given amount of practice (or change in conditions in which the task is performed) does not result in equal facilitation for the operation of all the cooperating factors contributing to individual differences in performance on the task.

The results of these studies also have implications for problems of aptitude measurement in the psychomotor area as well as to certain problems of criterion analysis. For a discussion of these problems the reader is referred elsewhere (**3**).

Summary

A practice task together with a number of reference tests was administered. A factor analysis of scores obtained at different stages of proficiency on the practice task together with scores on the reference tests was carried out. The results confirmed earlier findings with a different psychomotor practice task that considerable but systematic changes in factor structure occur as a function of practice. However, in contrast to the previous study, the present task did not become progressively less complex (in terms of the number of factors measured) as practice was continued. As before, there was an increase in the contribution of a factor common only to the practice task, but this increase was not as marked as with the previous task.

References

1. Adams, J. A. The prediction of performance at advanced stages of training on a complex psychomotor task. *USAF, Hum. Resour. Res. Cent., Res. Bul.*, 1953, No. 53-49.
2. Fleishman, E. A. Dimensional analysis of psychomotor abilities. *J. exp. Psychol.*, 1954, **48**, 437–454.
3. Fleishman, E. A., & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.
4. Fleishman, E. A., & Hempel, W. E., Jr. Factorial analysis of complex psychomotor performance. *USAF, Personnel & Training Res. Cent., Res. Bul.*, 1954, No. 54-12.
5. Gragg, D. B., & Gordon, M. A. Validity of the Airman Classification Battery AC-1. *USAF, Hum. Resour. Res. Cent., Res. Bul.*, 1950, No. 50-3.
6. Guilford, J. P., & Lacey, J. I. (Eds.) Printed classification tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 5.)
7. Lewis, D., McAllister, D. E., & Bechtoldt, H. P. Correlational analysis of the learning and relearning of four different tasks on the Modified Mashburn Apparatus. *J. Psychol.*, 1953, **36**, 83–109.
8. Lewis, D., McAllister, D. E., & Bechtoldt, H. P. Correlational study of performance during successive phases of practice on the standard and revised tasks on the SAM Complex Coordination. *J. Psychol.*, 1953, **36**, 111–126.
9. Melton, A. W. (Ed.) Apparatus tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)
10. Purdue Pegboard. (Examiner Manual.) Science Research Associates, 228 S. Wabash, Chicago, Illinois.
11. Reynolds, B. The effect of learning on the predictability of psychomotor performance. *J. exp. Psychol.*, 1952, **44**, 189–198.
12. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univ. of Chicago Press, 1947.
13. Woodrow, H. Factors in improvement with practice. *J. Psychol.*, 1939, **7**, 55–70.
14. Woodrow, H. The ability to learn. *Psychol. Rev.*, 1946, **53**, 147–158.
15. Zimmerman, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51–55.

(Received August 3, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

SOME EFFECTS ON PERFORMANCE OF CHANGES IN PERCEPTUAL LOAD[1]

R. CONRAD

Applied Psychology Research Unit, Cambridge, England

A common feature of skills of daily life is that signals requiring action if the skill is to be maintained may occur from more than one source, and more signals than one may occur simultaneously. This characteristic of the display side of skills has been discussed by the writer (**3**) under the term "load." Load can be given a quantitative definition as the number of (subjectively) independent streams of signals comprising a display, and this clearly refers to the perceptual content of the display.

Latterly, several studies have used tasks which specifically required *S* to consider a number of signal sources. They fall into three main groups: (*a*) Some are highly complex continuous tasks of a standard nature used primarily for the study of fatigue, no systematic attempt being made to vary the load content (**7, 10**). (*b*) A second class are multi-choice reaction time experiments, concerned mainly with relating reaction time to number of degrees of choice. In these studies the number of signal sources is the main experimental variable (**9**). The task, however, is designed so that a second signal cannot occur until *S* has responded to the first, and signals cannot be anticipated except as an error. (*c*) Finally, there are tasks which attempt to combine the complexity and seriality of the first group, with experimental variation of the number of signal sources (**3, 11**). The present paper describes a study falling into this latter group.

Both the above mentioned papers by Conrad (**3**) and by Mackworth (**11**) reported deterioration in performance with an increase in load, in the form of an increase in the number of errors. Mackworth postulated three main reasons for the increase in errors: first, because of the likelihood that with increased load there would be a greater probability of several signals occurring within a short time interval (crisis); second, because of increased difficulty of predetermining the order in which to deal with signals (disorder); and third, because of the presence of irrelevant information in some of the channels.

This study aims at determining whether *changes* in performance which accompany *load* changes can be expressed in terms of general validity. This implies less concern with the resultant of these changes in the form of an end score, and more, with the behavior underlying this end score. It was accordingly hoped to examine and extend Mackworth's postulates, with the addition of quantitative support.

METHOD

The apparatus used has been described in detail (**2**). Briefly it consists of a number of dials each carrying a revolving pointer and marks at the 6 o'clock and 12 o'clock positions. Mounted on a control panel in front of *S* are spring-loaded switches in spatial correpondence with the dials—one switch to each dial. If no control over the display is exercised by S, the pointers will stop at the first mark. If the appropriate switch is depressed as a pointer approaches a mark, and held down, the pointer will cross the mark with-

[1] The writer is indebted to Dr. N. H. Mackworth who read the manuscript in draft, and to Miss Barbara A. Hille for helping with the analysis of records.

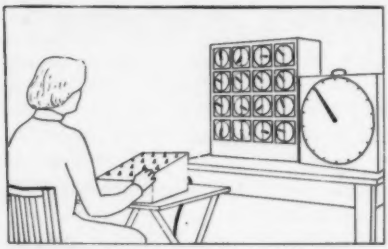


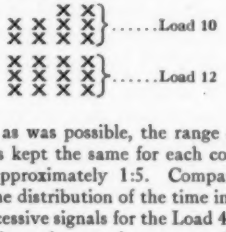
FIG. 1. Sketch showing lay-out of equipment.

out stopping and continue revolving. The *S*, who is seated 5 ft. from a 2 ft. wide display, is instructed to try to prevent pointers from stopping, and when stops occur to restart as soon as possible.

A large clock with a single hand provides a continuous record of success in achieving this aim. When a pointer is prevented from stopping, the clock hand moves forward one division; when a pointer stops, the clock hand moves back one division. During the time that a pointer is stopped, clock marks are lost at a rate proportional to the running speed of the stopped pointer (see Fig. 1).

Each pointer revolves at a controlled speed, but the speed of each is different from all the others. Thus, although the interval between signals is highly irregular, subject only to sampling errors, the display problem tends to be the same for each *S*. Cox and Smith (6) have shown that the nature of the distribution depends on the mean interval, the number of sources, and the relative signal frequency from each source. So long as the periods of each signal source are mutually irrational, the greater the number of sources, the closer to the exponential will the pooled distribution be.

In the present experiment, the mean signal speed was kept constant at 25 signals/min. The load was varied from 4 through 12 pointers in steps of 2. Spatially the display was built up as follows, each cross representing the position of a dial:



As nearly as was possible, the range of pointer speeds was kept the same for each condition of load at approximately 1:5. Comparison was made of the distribution of the time interval between successive signals for the Load 4 and Load 12 conditions (expected greatest difference). Although there were more very short and very long intervals for Load 12, the difference between

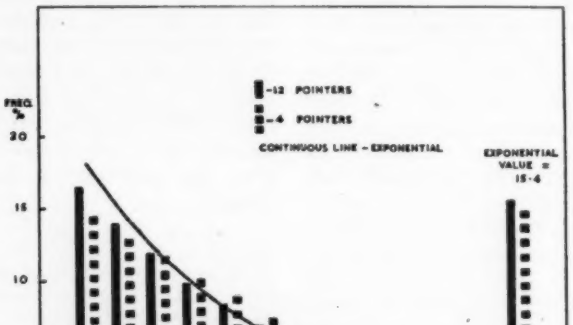


FIG. 2. Distribution of time interval between successive signals.

the two distributions was not statistically significant even at the .2 level.

These two distributions, together with the exponential for a mean speed of 25 signals/min., are shown in Fig. 2.

Success at this task depends on the ability to decide in advance which signal source will next require attention. Although the mean time provided for this decision was kept constant at 2.4 sec., the time available for each individual decision varies continuously. The perceptual data necessary for making the decision is always visible, and the main problem for *S* is to discover and use it in time to make the appropriate response.

In order to ensure a standard procedure, *S*s were instructed to use only one hand, and to press only one switch at a time. The average duration of a response was, after practice, about 1 sec.—.5 sec. on either side of the signal mark. But the shortest interval that could be tolerated between two successive signals if both were to be correctly dealt with was about .5 sec.

Twelve naval ratings were given 20 min. of practice at the task on each of five successive weekdays. The practice sessions were with a display of 16 dials at a speed of 15 signals/min. During the following week the experimental conditions were given in random order, each test lasting for 10 min.

Performance was recorded by a multi-pen recorder on paper moving at a speed of 1 cm. sec. The following events were recorded, separately for each dial but on a common time base: (*a*) the moment at which a pointer stopped; (*b*) the duration of such stops; (*c*) the moment at which a pointer crossed the target mark if a correct response was made; (*d*) the duration of responses. It was thus possible to build up a continuous accurate serial account in terms of time, of response in relation to events occurring on the display.

Results

For the five load conditions, the number of occasions per unit time *S* permitted a pointer to stop, and the mean duration of these stops are shown in Fig. 3. In both cases, the difference between the mean scores for any load interval of four dials is statistically significant at better than the .001 level. But the usefulness of these measures is largely limited to the function of distinguishing between test conditions on this particular task. Generalizations from them about the

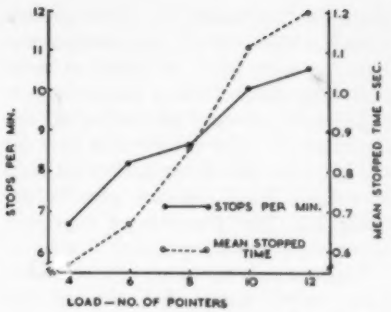


FIG. 3. Effect of load increases on performance.

effect of load changes in skill as a form of human behavior would be bounded by the extent to which one could justify this type of response as characteristic of skill in general.

A more detailed study of performance in relation to moment to moment variations in display conditions was made accordingly with a view to answering the following main question: under what temporal conditions did pointers stop, and were the conditions affected by load changes?

Of Mackworth's three postulates referred to earlier, by which he sought to explain the rise in error score with load increases, the presence of irrelevant information is not likely to be of importance in the present case. The visual field is structurally relatively simple; the action points are clearly defined in each case; the perceptual information conveyed by a signal is easily interpreted. The nearer to the target mark that a pointer approaches, the more relevant is the information conveyed. But strictly speaking, at no point is the information irrelevant, merely, more or less relevant. There is never present in the display, information which *S* ought to disregard. But Mackworth's other two factors need to be considered.

Crisis.—With continuously changing displays in which signals occur at irregular time intervals, there may be long intervals when no signal occurs, followed by the occurrence of several signals bunched together. If more than one signal source is involved,

there is the possibility that several signals will occur simultaneously. Error is more likely to occur at these bunches, than when a signal is preceded and followed by a long signalless period. The problem is how far the observed error increase which accompanies load increase can be explained by the presence of a greater number of bunches per unit of time.

There are two cases to consider. First, when the distribution of the time interval between successive signals is strictly random, and second, when it is not random, is irregular, but is known. In the former case of the exponential distribution, the probability with which m events will occur in t time depends only on the expected number of events in t time, that is, on the mean signal speed. The number of signal sources is irrelevant. If the mean frequency with which signals occurred was constant, there would be no more bunches of signals with high loads than with low. In such a case, although "crises" would occur, they would occur with the same probability irrespective of the load. One would expect errors to occur at such crises, but one could not expect the number of crisis errors to change with load changes.

Both with the present task and that used by Mackworth, the interval distribution was not strictly random, but was based on a number of slightly different periodic sources. Signals could occur simultaneously, the maximum number being the number of sources comprising the display—the load. The longest interval would be equal to the shortest period. But with more signal sources, the distribution approaches closer to the exponential, i.e., there are more very short intervals, and also more very long. In the present context this means that the probability of m events occurring in t time increases as the load increases so long as t is short. Hence one might expect an increase in error from this factor.

TABLE 1

THE NUMBER OF BUNCHES PER MINUTE

m	Load = 4			Load = 12		
	t(Sec.)			t(Sec.)		
	1.0	1.5	2.0	1.0	1.5	2.0
3	.64	1.00	1.70	1.00	1.78	2.92
4	.00	.06	.18	.06	.24	.46

A difficulty lies in defining m and t. With a mean signal speed of 25 signals/min. one event is expected in 2.4 sec. How many events constitute a bunch, and how close together should they occur? Table 1 shows for the two extreme load values of 4 and 12, the frequency per minute with which bunches occur for m values of 3 and 4, and t values of 1.0, 1.5 and 2.0 sec.

The actual number of errors (stopped pointers) per minute with Loads 4 and 12 was 6.7 and 10.6, respectively. If one takes the most favorable view and considers a case when every signal in a bunch leads to error, even with a t value of 2.0 sec., not enough errors would arise from this factor to account for the observed number. It is clear on the one hand that "crises" are a source of error which will increase as load is increased in certain types of skill, but it is also clear that there are other sources of error. In general, the relative importance of this factor depends primarily on the degree to which the distribution of the time interval between signals is modified by load changes. With skills in which the display changes are, as it were, artificial, such as industrial machine-

minding skills, this factor will be important (1). With more natural displays such as traffic situations, it will be less important.

Disorder.—This type of error arises when *S* fails to determine the correct order in which signals occur. With the present task, misjudgment of this kind inevitably leads to at least one stopped pointer and sometimes two. In misjudging, *S* might attend first to the second of two signals, making a correct response to it, while the first signal through neglect leads to an error. Or quite frequently the record shows that *S* began a response to the second signal, changed his mind before the signal occurred, returned to the first correct signal, but too late to prevent the pointer from stopping, returned then to the second signal, and was too late for that also. In this way two errors occur.

There is no doubt that this kind of error also contributes to the gross error score for all loads. Further, it occurs more often with high loads than with low. Comparing the 4- with the 12-pointer condition, with the same number of signals presented per minute, the number of occasions when disorder can be clearly detected was .92 and 2.55 per min., respectively. This led to 1.26 and 3.65 errors per minute. There are not significantly more double errors with the greater load.

A further point was investigated. How far apart do two successive signals have to be for disorder to be unlikely, and is this interval affected by load changes? The records of 6 *Ss* were analyzed for Loads 4 and 12. The mean interval at which disorder occurred was .41 and 1.19 sec., respectively. In each case the range was 0–2.9 and 0–4.7 sec. The longest possible interval between two signals was 7.7 sec. for Load 4 and 14.2 sec. for Load 12. Thus only the shorter intervals in the distribution were liable to this error.

It should be noted that a good many of the disorder errors which were found occurred within bunches of signals, where short intervals between successive signals were likely. In determining the relative importance of the different sources of error, disorder errors cannot always be distinguished from those due to crisis. One recorded error may be associated with several known sources of error.

There is an ambiguity in the significance of this type of error which needs clarification. In keeping the signal speed constant and increasing the load from 4 through 12, the average pointer speed in rpm is reduced approximately by a multiple of three. It will be noted that the mean size of interval when disorder errors occurred was approximately three times as great with Load 12 than with Load 4. This suggests that in the present case disorder errors did not arise through "load stress," but through difficulties in making certain spatial discriminations. The greater the load, the slower the pointers moved; the slower the pointers moved, the shorter the linear interval to the next target mark for equal time intervals.

Since it is not possible with the present apparatus to vary the load without changing the pointer speed, a direct test of the effect on disorder errors of pointer speed was not possible. But if the load was kept constant and pointer speed increased, although signal speed would be increased, proportionately, fewer disorder errors should occur if pointer speed was a major factor contributing to their incidence. Accordingly an experiment was carried out using six well-practiced naval ratings as *Ss*, in which each *S* did two tests of 10 min.

each, one at 15 signals/min., the other at 30 signals/min. in alternate order. In both cases Load 8 was used, and, of course, doubling the signal speed doubled the pointer speed.

Simple comparison of the mean size and the frequency of disorder errors was not possible because, unlike the comparison of the two load conditions, the frequency distribution of the interval between signals was very different for the two speed conditions. In the former case, an interval of given size occurred with approximately equal frequency. This was not true in the latter case (**5**). The actual distribution could, however, be calculated (**6**), while the distribution of the interval when disorder errors occurred, was measured from the records of performance. Thus the probability of disorder error for any given size of interval could be obtained for the two pointer-speed conditions. For the intervals .5, 1.0, 1.5, and 2.0 sec., these are shown in Table 2.

It is clear that for any given interval, slowing the speed of the pointers does not increase the chance of a disorder error. The apparent decrease at slow speed is in no case statistically significant at the 5% level of probability.

Sequence.—There is another source of error likely to be present in all skills. This is insufficient time between two successive signals, either to determine which is the next signal, or having determined it, to make an adequate response. In connection with a somewhat different skilled task the writer (**4**) reported that the effectiveness of a response depended to some extent not only on recent events, but also on the proximity of a subsequent event. It was further suggested that independently of the absolute time values involved, this balance might be affected by the average time available in the signal-response series as a whole —that is, the mean signal speed.

TABLE 2

PROBABILITY OF DISORDER ERROR IN RELATION TO THE INTERVAL BETWEEN SUCCESSIVE SIGNALS FOR SLOW AND FAST POINTER SPEEDS

Pointer Speed	*N*	Interval Between Signals (Sec.)			
		0–.4	.5–.9	1.0 –1.4	1.5 –1.9
15 signals/min.	900	.26	.25	.14	.10
30 signals/min.	1800	.31	.33	.25	.26

In the present case the mean signal speed was constant at 25 signals/min. But in varying the load, the attentional spread changes, without a change in the average time available for it. Again, therefore, there is the double question. To what extent are errors due to shortage of time, and how far do load changes affect the time which is necessary for adequate response?

A temporal sequence of events studied in earlier work was again selected. The sequence consisted of: (*a*) the end of Response 1; (*b*) the occurrence of Signal 2; (*c*) the occurrence of Signal 3. The two intervals were called R_1S_2 and S_2S_3, respectively, and the adequacy of the response to Signal 2 was studied in relation to the duration of these two intervals, one before and one after the signal. All of these three events were recorded, and the analysis was made with Loads 4 and 12, using 1,850 signals in each case.

In making the analysis, measurement was in units of .25 sec. up to an interval size of 1 sec. For intervals greater than this, units of .5 sec. were used. Intervals greater than 3 sec. were not distinguished. For convenience the data are expressed in

TABLE 3

FREQUENCY (%) WITH WHICH A SIGNAL IS CORRECTLY DEALT WITH AS A FUNCTION OF THE TEMPORAL STRUCTURE OF THE SEQUENCE WITHIN WHICH IT OCCURS

S_2S_3 (Sec.)	R_1S_2 (Sec.)			
	0–.9	1.0–1.9	2.0–2.9	3+
Load 4				
0– .9	40.0	70.0	72.5	68.3
1.0–1.9	54.3	87.8	87.3	88.9
2.0–2.9	60.9	92.5	83.3	87.0
3+	46.5	86.5	94.0	80.7
Load 12				
0– .9	24.7	50.8	57.3	47.9
1.0–1.9	28.1	63.3	72.1	66.4
2.0–2.9	25.0	67.3	77.8	76.5
3+	31.9	67.1	73.7	81.8

Table 3 in units of 1.0 sec. which show most clearly the general trends. This table gives the percentage number of times that a signal was responded to correctly when it occurred in sequences of varying temporal structure.

In general it is evident from Table 3 that: (*a*) the longer the "free" period before a signal, the greater the probability that it will be successfully dealt with; (*b*) independently of this prior interval, the longer the interval until the next signal, the greater is the probability that a signal will be successfully dealt with; (*c*) for a sequence of given temporal structure, the smaller the load, the more likely the signal will be successfully dealt with. This particular error indication, therefore, does not depend primarily on the nature of the distribution of the time interval between signals, but on the number of signal sources. And it would appear to be so even if the sequence frequencies were absolutely identical.

It should be mentioned that the apparent reversal of trend for the very largest interval is quite misleading. It arises in the table entirely as an artifact of the scoring method and does not reflect true behavior. The occurrence of a signal is an event which theoretically is without duration, while correct responses have finite duration. If one signal occurs during the response to another, it will tend, on the scoring method adopted, to have an artificially long R_1S_2 or S_2S_3 period depending on whether it occurs just before or just after the other. The present method was used because it seemed to be the most foolproof of those considered which were psychologically justifiable. The cases which give artificial values are clearly defined on the performance record, and no subjective interpretation is necessary.

Runs.—Discussion with *S*s led to the suggestion that recovery from an error was more difficult with greater loads. A method of checking this impression was to measure the length of runs of successive signals leading to error, both in terms of time and of number of signals in the run. A run for this purpose was defined as an interval from the first signal not correctly dealt with, to the next signal correctly dealt with. Table 4 shows the result of this analysis.

TABLE 4

RUNS: THE EFFECT OF LOAD INCREASE ON RECOVERY FROM AN ERROR

Measure	Load 4		Load 12	
	Mean	*SD*	Mean	*SD*
Mean number of signals per run of errors	1.58	1.04	1.85	1.35
Mean duration of runs (sec.)	2.65	1.93	3.14	2.47
Mean interval between signals in runs (sec.)	1.68		1.70	

The difference between the means of the two load conditions for the length of run using either measure is statistically significant at the .05 level. The signal density in runs is not significantly different in Loads 4 and 12. Thus, although for this measure the objective situation in terms of time was not more difficult with the greater load condition, recovery from error appears to have been less easy.

Distribution of attention.—The 12-pointer task provided a convenient opportunity to observe whether in this type of situation attention was distributed over the display strictly according to objective display demands. Alternative possibilities were that certain spatial positions (e.g., top left-hand) would absorb a disproportionate amount of attention, or that sources of relatively high signal frequency would receive too much attention. As an index of attention the number of correct responses for each dial was taken. An expected number was calculated from the objective signal frequencies. In no single case were the observed and expected numbers significantly different.

Discussion

In trying to understand the nature of human skill and the factors which disturb it, it is clearly desirable to get as close to the performance being studied as is compatible with clarity of observation and accuracy of recording what is observed. The results of this experiment illustrate at least two different degrees of closeness to *S*s' performance. It would have been quite simple and legitimate merely to record on counters the number of correct responses and the number of pointer stops and then calculate the mean stopped time. If this were done, conventional results of the effect of load changes would have been accurately obtained without recourse to any other observation of the behavior which resulted in the final figures on the score counters.

The more detailed picture is useful because it suggests some of the ways in which behavior was modified when the load was increased; the modification leading to the end score. It is not suggested that the picture is complete. There may be modifications which have not been detected. The terms in which the records were examined derive partly from other studies (**4, 5, 11**), partly from subjective impressions, and partly from *E*'s observations of *S*s working. Not all of these impressions and observations were correct. There was no particular tendency with high loads to concentrate attention only on certain signal sources. On the other hand, the strong and uncomfortable feeling that one was more "put off one's stride" by an error when the load was high than when it was low appears to have been justified by the quantitative analysis of runs.

In all cases the detailed criteria of performance are such that they would be valid when applied to any skill other than simple tracking and in which *S* did not have complete control over the temporal structure of the stimulus series. More crude criteria have not this quality. They depend too much on the specialized character of the skill.

In this connection it should be noted that there was no breakdown of skill in a massive sense. Even under the greatest load condition, pointers were kept moving (the object of the skill) for about 98% of the total time. When breakdown did occur, in the sense of loss of control over the environment, recovery was relatively rapid. The duration of runs is an indication of the duration of loss of control. With greatly increased stress one would predict that runs will extend over such durations as to justify a conclusion of complete perceptual breakdown, i.e., no judgment correctly made until it is too late to translate it into adequate action.

In a previous paper (**5**) the writer outlined a concept of speed stress. This was considered to be a psychological reaction on the part of *S* to display conditions. It had the effect of worsening performance beyond that which might be expected from the physical characteristics of the display. Although it would be in-

cautious in the present case to assert that the deterioration in performance with increasing load was partly an effect of load stress analogous to that of speed stress, the idea is worth consideration.

Clearly, increasing the load from 4 through 12 signal sources does change the character to the display so that more time might be needed to make the necessary judgments simply because of the greater scanning coverage required. This is an obvious explanation of the results. But it involves a good many assumptions.

It is quite certain that *S*s do not need to fixate visually every signal source before making a decision. If they did, according to the data on eye fixation time of Fitts (8), about 6 sec. would be required for each decision on the 12-pointer task. Table 3 eliminates this even as a remote possibility. Senders (12), on the other hand, reports that the time required to detect 1 misaligned pointer among 16, when alignment is always at the 9 o'clock position, is .25 sec. The eye movements which might be necessary in the present experiment are more likely to resemble those of Senders' *S*s, since near-alignments are being sought rather than the making of quantitative readings of an instrument panel as in the Fitts' study. But direct eye-movement studies are unlikely to be of help. The visual angle subtended by the entire display was only about 25°. A fairly accurate idea, therefore, of the location of all the pointers would be quite possible without recourse to visual scanning at all. Visual searching does occur, and eye fixation is necessary to estimate which signal source will next require attention. But it is doubtful whether the results can be explained in terms of the temporal or spatial pattern of visual behavior.

Alternatively it could be argued, that with higher loads, comparisons of the speed and position of more pointers will be needed per response, in the same average amount of time that is available with low loads. Again this is true to some extent. But the analysis in the results of errors due to crisis shows that the number of occasions when this *must* happen is not sufficient to account for the number of errors. It might in fact be argued that at all loads one glance discards all but one or two sources as being likely to provide the next signal. In which case the numerical value of the load would be irrelevant.

Thus a concept of load stress analogous to that of speed stress remains a worthwhile hypothesis which could partly account for the observed changes in behavior. This hypothesis implies that subjects confronted by a task which apparently demands more work in the same amount of time actually require longer for identical items of behavior. That is, there is a genuine change in thresholds of behavior brought about by an inaccurate assessment of the nature of the physical environment.

Summary

An experiment was carried out using 12 practiced *S*s in which the perceptual load was systematically varied. The aim of the experiment was to determine whether changes in serial performance as well as changes in score could be associated with the experimental variable.

Increasing the number of pointers from 4 through 12 had the effect of increasing both the number and the duration of errors. Detailed performance records were scrutinized clinically, and a number of features thus isolated were subjected to conventional statistical treatment. It was found that:

1. Errors occurred when signals requiring action were bunched. It was pointed out that with load changes the errors arising from this source would change in certain skills, but not in all.

2. Errors occurred when *S*s failed to determine the correct signal order. These errors occurred only when the interval between successive signals was short. Increasing the load had the effect of increasing the frequency of such errors, and of increasing the maximum interval between two signals likely to result in such error. A subsidiary experiment showed an association between the incidence of this type of error and a general shortage of time.

3. With load increases, errors were more likely to arise in event sequences of identical temporal structure, even when an essential element of the sequence was an event which occurred after the error.

4. Recovery from error took longer with greater loads.

5. Even under the condition of greatest stress, attention was apparently distributed strictly according to display demands rather than in accord with either a spatial or a signal-frequency pattern.

The results are discussed briefly in terms of their value as general criteria of changes in skilled performance under stress. A concept of load stress analogous to that of speed stress is postulated, and certain objections made from a consideration of eye movements are reviewed.

References

1. Benson, F., & Cox, D. R. The productivity of machines requiring attention at random intervals. *J. Roy. statist. Soc.* 1951, **13**, 65–82.
2. Carpenter, A., & Conrad, R. An apparatus providing signals at "random" intervals. *Quart. J. exp. Psychol.*, 1953, **5**, 87–88.
3. Conrad, R. Speed and load stress in a sensori-motor skill. *Brit. J. industr. Med.*, 1951, **8**, 1–7.
4. Conrad, R. The temporal accuracy of response in a sensori-motor skill. *Med. Res. Coun., Appl. Psychol. Unit*, 1953, Rep. No. 186.
5. Conrad, R. Missed signals in a sensori-motor skill. *J. exp. Psychol.*, 1954, **48**, 1–9.
6. Cox, D. R., & Smith, W. L. The superposition of several strictly periodic sequences of events. *Biometrika*, 1953, **40**, 1–11.
7. Davis, D. R. The disorganization of behaviour in fatigue. *J. Neurol. Neurosurg. Psychiat.*, 1946, **9**, 23–29.
8. Fitts, P. M., Jones, R. E., & Milton, J. L. Eye-movements of aircraft pilots during instrument landing approaches. *Aero Engng. Rev.*, 1950, **9**, 1–16.
9. Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.
10. Hoffman, A. C., & Mead, L. C. The performance of trained subjects on a complex task of four hours' duration. *OSRD* Report 1701, 1943, Tufts Coll., Medford, Mass.
11. Mackworth, N. H. Some recent studies of human stress from a marine and naval viewpoint. *Trans. Inst. marine Engng.*, 1952, **64**, 1–10.
12. Senders, V. L. The effect of number of dials on qualitative reading of a multiple dial panel. *USAF, WADC, Tech. Rep.*, 1952, No. 182.

(Received July 15, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

ABSOLUTE JUDGMENTS AS A FUNCTION OF STIMULUS RANGE AND NUMBER OF STIMULUS AND RESPONSE CATEGORIES [1]

CHARLES W. ERIKSEN AND HAROLD W. HAKE

The Johns Hopkins University

The method of absolute judgment, or as it is sometimes called, the method of single stimuli, may be briefly characterized as a type of judgment in which *O* is required to identify by a name, number, or value each member of a set of individually presented stimuli. Since the method does not provide any explicit standard for comparison, *O* is required to base his judgments upon some subjective standard or reference point.

Although the method is very similar to the situations under which people are required to make many important judgments during their daily activities, the method has been relatively neglected both in the area of psychophysics and as a means of studying cognitive behavior. In an early study Wever and Zener (**13**) concluded that the absolute method yielded discrimination functions comparable to those obtained by other psychophysical methods. Recently there has been a new interest in the method owing to its amenability to information analysis which has provided a means for stating the amount of information which can be transmitted by judgments along psychological dimensions, or, equivalently, the number of different stimulus values along such dimensions which can be judged or identified absolutely with little or no error (**4**). This analysis has been applied to the dimensions of loudness (**3**), pitch (**10**), and hue (**1**).

In the present experiment the informational measure of absolute judgmental accuracy has been used primarily as a useful means of studying the judgmental process itself. We were concerned with the extent to which the number of absolutely discriminable stimulus categories can be affected by subjective anchoring effects associated with the range and density of the stimulus dimension and with the number of response categories available to *S*s for expressing discriminations or judgments.

Previous research has indicated that absolute judgments tend to be anchored by the range and density of stimuli presented for judgment (**6, 11, 12**), and it has been suggested that such judgments are affected by the entire range of stimuli presented (**7**). The effects of drastic variations in auditory stimulus factors upon the information measure have been found to be slight, however (**10**). The present study provides more detailed information about the effects of variation in both stimulus range and response factors upon the information measure of the accuracy of absolute judgments of visual size.

METHOD

Design.—The size dimension, consisting of squares varying in size from 2 to 82 mm. sq., was the stimulus dimension used in the present study. Two size ranges were studied: one where the size of the squares was spread over the range 2–42 mm. sq. and the second where the squares

[1] This research was performed under Contract AF 33(038)-22642 between the Aero Medical Laboratory, Wright Air Development Center, and The Johns Hopkins University.

covered the range 2–82 mm. sq. For the response variable, 5, 11, or 21 response categories were permitted. A similar breakdown was employed for the number of stimulus categories. Discrimination was studied when 5, 11, and 21 different sized squares were presented for judgment. With five stimulus categories, the five squares increased in size by 10 mm. per side for the 2–42 range and 20 mm. per side for the 2–82 mm. sq. range. With 11 categories the increment in size was 4 mm. per side for the 2–42 and 8 mm. per side for the 2–82 mm. sq. range, and with 21 categories the increment was 2 and 4 mm. per side, respectively. Irrespective of the number of stimulus categories the smallest sized square presented for judgment was always 2 mm. sq. and the largest was always 42 mm. sq. when the restricted size range was used and 82 mm. sq. when the extended size range was used.

The three experimental variables, size range, number of responses, and number of stimuli, were arranged in a three-dimensional factorial design and five *O*s were randomly assigned to each of the 18 treatment conditions.

Experimental procedure.—The *O*s were told that this was an experiment to determine how well people could distinguish between different sized squares. They were instructed to call the "smallest" square by the number "1" and the "largest" square "5," "11," or "21," depending upon the number of responses the particular *O* was to be permitted to use. The *O* was further told to assign the other response categories to the squares in order of increasing size.

Each *O* was given a practice session, consisting of two presentations of each stimulus category, to familiarize him with the task and the stimuli he was to be judging. In this practice session the same number of stimulus and response categories were used as were to be used in the experimental session with *O* and he was informed after each judgment whether the judgment was right or wrong. Since for some *O*s the number of stimulus categories exceeded the number of responses and for other *O*s the situation was reversed, it was necessary to code or group the responses (or stimuli) in order to be able to label a given judgment right or wrong. In coding, the responses were equally divided among the stimuli, or the stimuli were equally divided among the responses, whichever case obtained. Thus, in the case where *O* was required to judge 5 stimuli using 11 responses, a response of "1" or "2" to the smallest stimulus was called "correct"; a response of "3" or "4' to' the next larger stimulus was called correct, etc. The extra response was assigned to the middle stimulus category. No information was given concerning the correctness of responses during the experimental series. The stimuli were presented to *O* one at a time in a random order. The random orders varied between *O*s within a treatment condition, but the same five orders were used in all 18 treatments.

Stimulus materials.—The stimuli were a series of 41 squares cut from a dark gray paper and pasted in the center of 8 × 11-in. white cardboards.

Subjects.—A total of 90 *O*s served in this experiment. They were male undergraduate students enrolled in elementary psychology courses at The Johns Hopkins University.

Statistical analysis.—The data have been analyzed in terms of information theory after the manner suggested by Garner and Hake (**4**) and McGill (**8**). Garner and Hake have shown that the measure of information transmitted (I_t) is interpretable as the maximum number of stimulus categories that can be perfectly discriminated within a given range on a stimulus continuum. Since I_t is measured in bits, it is necessary to take the antilogarithm to the base of two in order to arrive at the number of stimulus categories that would yield perfect discrimination. Thus, if a particular range of a stimulus dimension was found to transmit 2.00 bits of information, this would indicate that a maximum of four stimuli could be selected from this range to give errorless discrimination under the absolute method. In this paper we have defined discrimination in terms of I_t.

The choice of the information model for our analysis was also dictated by the fact it applies an inherent correction for chance performance. As an example, if we compared discrimination when *O*s used two response categories as opposed to four categories, our comparison would be confounded by the fact that *O*s would be expected to be correct half the time by chance using two categories, but chance performance with four categories would only be one-fourth. The I_t measure corrects for such chance effects.

Information transmitted may be expressed as the difference between the total uncertainty in *O*'s responses independent of stimuli and the average uncertainty of his response when we know which stimulus he is judging. Stated symbolically, $I_t = I_r - E_r$, in which I_r is a measure based only upon the response distribution of *O* independent of specific stimuli and E_r is an average of similar measures of the distributions of responses made to specific stimuli. If there is no discrimination, I_r equals E_r. If I_t is some positive value greater than zero, some discrimination is taking place, and if the number of response categories is increased, the increase in I_r is associated with a comparable increase in E_r, provided discrimination remains constant.

Defining I_t as the difference between I_r and E_r suggests an alternative interpretation of I_t in

terms of discrimination. It may be considered the amount of variation in *O*'s responses which is attributable to variation in the stimuli. Although analogous to conventional correlational analysis, in this sense, measures of information or uncertainty cannot be transformed directly into more familiar variance measures because the former responds to the "coding" relations between stimuli and responses rather than to amount of error (**4, 5, 8**).

Results

General effects.—The first step in our analysis was to estimate the overall statistical reliability of the results. Accordingly, the amount of information transmitted (I_t) by each of the five *O*s under each of the 18 experimental treatments was computed and analyzed by a three-way analysis of variance (length of stimulus dimension, number of stimulus categories, and number of response categories).[2] The results of this analysis show only two significant effects ($p \leq .05$). The first of these pertains to the length of the stimulus dimension. When the sizes of the squares are distributed over the range of 2 to 82 mm. sq., discrimination is better than when the size range is only 2 to 42 mm. sq. ($F = 12.8$, $p < .001$). The other significant effect is an interaction between the number of stimulus and the number of response categories ($F = 8.3$, $p < .001$). Neither number of stimulus nor number of response categories alone is a significant source of variance when evaluated by the significant interaction between them.

TABLE 1

Mean Values of I_t Obtained Under the Various Experimental Treatments

(The upper values in each cell are for the stimulus range 2–42 mm. sq. and the lower values are for the stimulus range 2–82 mm. sq. The middle values are the average for the two ranges.)

Number of Responses	Number of Stimuli		
	5	11	21
5	1.98 *2.08* 2.17	1.51 *1.65* 1.78	1.47 *1.49* 1.50
11	1.86 *1.93* 2.00	1.91 *2.07* 2.23	1.77 *1.90* 2.02
21	1.99 *2.03* 2.07	2.04 *2.14* 2.23	2.08 *2.08* 2.07

Note.—For comparing means within and between the two stimulus ranges, mean differences of .27 and .36 are significant at approximately the .05 and .01 levels, respectively. For the means from the combined ranges, differences of .19 and .25 are significant at the .05 and .01 levels. These significance levels are based upon the error term from the analysis of variance.

The interaction between the number of stimulus and response categories is apparent in Table 1. Here the mean values of I_t for the various combinations of stimulus and response categories are shown for both size ranges, 2 to 42 and 2 to 82 mm. sq., as well as the means obtained by combining the two size ranges. Discrimination is equally good for 5, 11, and 21 stimulus categories provided the number of responses permitted is equal to the number of stimuli. Discrimination is the same with 5-5 as it is with 11-11 and 21-21. This table can be well summarized by noting that most of the variation between the cells depends upon whether the

[2] Application of Bartlett's test failed to provide evidence for rejecting the null hypothesis that the variances were homogeneous. There was the additional consideration as to whether the I_t scores were normally distributed since they represent a log transform of the number of categories that could be discriminated without error. While there is no clear empirical evidence on this question, we considered it appropriate to perform an analysis of variance on these scores. Norton (**9**) has shown that moderate departures from normality have but little effect upon significance levels. Since in the present study all effects were highly significant ($p = .001$), or clearly insignificant, the uncertainty concerning the normality of the distribution of I_t for different individuals should not invalidate the analysis of variance.

number of responses permitted *O*s is less than the number of stimulus categories to be judged. When the responses are equal to or greater than the number of stimuli, discrimination tends to be equally good irrespective of the number of stimulus categories involved. But when there are fewer response categories than stimuli, discrimination is impaired.

Effects of number of response categories.—The finding that discrimination becomes poorer when the number of responses permitted is less than the number of stimuli poses some interesting problems. Since discrimination is equally as good for the 5-5 as with 11-11 or 21-21, it would seem that five response categories are sufficient to express the accuracy of *O*'s judgment. The counter finding merits further analysis.

As a first step in this analysis we need to determine what use *O*s made of the responses that were permitted them. Inspection of the data revealed that all *O*s tended to distribute their responses throughout the permitted range, but they did not use all responses equally often. In general, responses at the small end and the middle of the range were used considerably more often than responses at the large end of the range. In Fig. 1 the value of I_r shows the uncertainty, in bits, that exists in the average *O*'s responses. Here I_r has been plotted as a function of the number of responses permitted and the number of stimulus categories presented for judgment.

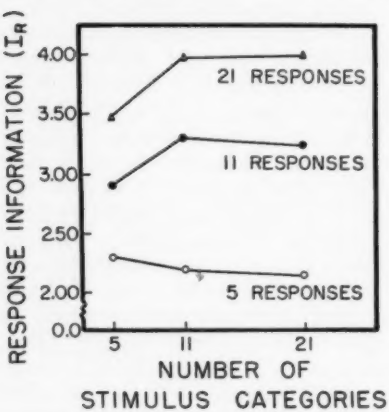


FIG. 1. Total response uncertainty (I_r) as a function of the number of stimulus and response categories. The data from the 2–42 and the 2–82 mm. sq. range have been averaged.

If *O*s had used all permitted responses equally often, I_r would equal 4.39 with 21 responses, 3.46 with 11, and 2.32 with 5 responses. In the case of 5 responses they closely approximated complete and equal usage. With 11 and 21 responses, the completeness with which *O*s use responses varies more markedly with the number of stimulus categories, but even here there is a strong tendency for all responses to be used. Expressed percentagewise and averaging through stimulus categories, 96% of the maximum uncertainty was used with 5 responses and 93% and 88% with 11 and 21 responses.

The marked reduction in I_r observed for 11 and 21 response categories when 5 stimulus categories are presented for judging is probably attributable to the fact that 5 categories were almost within *O*'s range of errorless discrimination. As was seen in Table 1, *O*s could transmit slightly more than 2 bits of information. This indicates that between four and five categories could be discriminated without error. With *O*s able to subjectively identify each category, there was less need for the surplus responses to express their indecision. The fact that *O*s used more responses than necessary is likely due to their interpreting the instructions as requiring that they make use of all response categories.

Error analysis.—Once we had determined that *O*s made a reasonably

complete usage of the responses available to them, the next step in our analysis was to determine the nature and distribution of response errors. As one measure of error we have used the uncertainty, in bits, of *O*'s response when we know which stimulus category he is judging. This measure of error (E_r) was computed for each stimulus category under all treatment conditions. In order to illustrate the effect upon E_r when the number of stimuli are held constant but the number of responses varied, we have presented the values of E_r for each stimulus category when 5, 11, and 21 responses were permitted and 5, 11, and 21 stimuli presented. These curves are shown in Fig. 2.

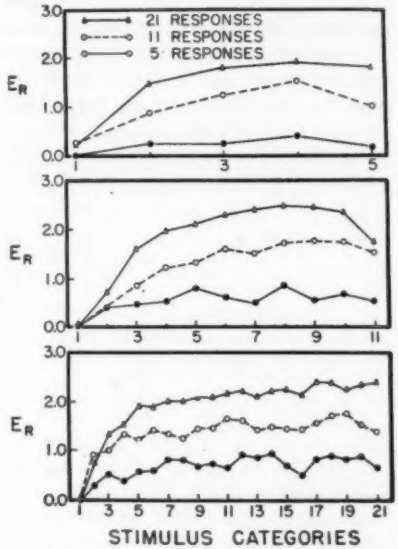


FIG. 2. Response equivocation per stimulus category (E_r) as a function of the number of stimulus and response categories. The data from the 2–42 and the 2–82 mm. sq. range have been averaged. The abscissa shows the stimulus categories in serial order from the smallest to the largest stimulus. The upper box shows the curves obtained for different numbers of responses when 5 stimulus categories were used. The middle and lower boxes are for 11 and 21 stimulus categories, respectively.

If we consider first the E_r curves when 21 stimuli were used, we can see the reason why discrimination is poorer when the number of responses is less than the number of stimuli to be judged. For discrimination to remain constant with a given number of stimuli, the average value of E_r must vary in a direct one-to-one relation as we vary I_r (the log to the base 2 of the number of available responses). If total response uncertainty (I_r) is decreased by permitting fewer responses, the average value of E_r must decrease by the same amount for I_t to remain constant. As can be seen from Fig. 2 this condition did not prevail.

When with 21 stimuli the number of response categories is reduced from 21 to 11 or 5, the uncertainty of response per stimulus shows a proportionate decrease only in the middle and upper range of stimulus values. The response uncertainty for the smaller squares remains unchanged or decreases only slightly. From Fig. 1 we can see that the average reduction of E_r should have been about .75 bits when the number of responses was reduced to 11 and about 1.8 bits when reduced to only 5. These reductions in E_r were closely approximated from Stimulus 5 through 21, but the reduction in E_r for the three lowest stimuli was considerably less than the expected amount. Since a similar effect can be noted for 11 stimuli with 5 responses, it appears that the failure to obtain a reduction in the average E_r equal to the experimentally manipulated reduction in I_r is due to the end stimuli which fail to show a proportionate reduction in equivocation when the number of response categories is reduced.

When the number of responses is greater than the number of stimuli, on the other hand, discrimination tends to remain constant. The curves for 5

TABLE 2

PERCENTAGE OF RESPONSES LARGER AND SMALLER THAN THE MODAL RESPONSE TO EACH STIMULUS WHEN 11 STIMULI AND 11 RESPONSES WERE USED

Response Classification	Stimuli										
	1	2	3	4	5	6	7	8	9	10	11
2–42 Mm. Range											
Smaller		1	11	25	9	31	31	41	45	40	62
Modal	1	2	3	4	5	6	7	8	9		11
Larger	1	28	26	27	52	38	34	28	29	23	
2–82 Mm. Range											
Smaller		0	5	22	33	41	19	22	24	52	38
Modal	1	2	3	4	5	6	6	7	8	10	11
Larger	0	21	36	34	25	24	39	38	46	6	

stimulus categories and for 11 stimulus categories with 21 responses in Fig. 2 show the adjustment and distribution of errors that accompany this situation. There is a slight tendency among the low stimulus categories for E_r to lag behind the increases in I_r, obtained by increasing the number of permitted responses, but this tendency is compensated by *O*s' using their responses with a little more abandon in the middle and upper range of stimuli. As a result the average gain in E_r closely approximates the increase in total response uncertainty.

There appear to be several factors involved in determining this variation in E_r as a function of the number of response categories. One pertains to the nonsymmetrical distribution of errors obtained for the stimuli on the low end of the range. Not only were errors fewer for these stimuli, as can be seen in Fig. 2, but when errors occurred, they were mainly one directional—that is, the errors were made by assigning to these stimuli response numbers that were too large. The extent to which this occurred can be seen in Tables 2 and 3 where we have presented the modal response to each stimulus and the percentage of responses larger and smaller than the mode for the 11-11 and the 21-21 conditions. As can be seen from these tables, it is not until the fourth stimulus that the errors began to approximate a symmetrical distribution. For the first three stimuli nearly all the errors are made as a result of call-

TABLE 3

PERCENTAGE OF RESPONSES LARGER AND SMALLER THAN THE MODAL RESPONSE TO EACH STIMULUS WHEN 21 STIMULI AND 21 RESPONSES WERE USED

Response Classification	Stimuli																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2–42 Mm. Range																					
Smaller		3	5	12	41	11	28	59	36	31	19	61	46	25	19	29	37	24	10	32	38
Modal	1	2	3	4	5	5	6	8	8	8	10	10	10	10	10	11	12	12	12	14	15
Larger	1	13	42	24	23	52	42	13	27	41	52	9	14	41	51	40	32	41	57	41	39
2–82 Mm. Range																					
Smaller		1	4	26	27	51	43	67	56	39	51	24	35	26	46	23	52	60	37	24	80
Modal	1	2	3	5	6	8	8	10	10	10	11	10	11	11	13	12	15	16	15	15	21
Larger	0	31	44	35	43	17	34	11	23	35	25	56	44	53	33	55	29	22	46	61	

ing these stimuli by larger response numbers. For the middle and upper range of stimuli, errors are more bidirectional in that for a given stimulus errors are about as apt to result from calling the stimulus by too small a response as by too large a response.

These peculiarities in the distribution of errors work to a disadvantage when there are less response categories than stimuli. When *O* has to group stimuli under a common response, for example Stimuli 1 and 2 under the common Response 1, there is little saving in errors since Stimulus 2 had hardly ever been confused with Stimulus 1. On the other hand, in the middle of the stimulus series, Stimuli 5 and 6 tend to be confused with each other equally often and thus, when these stimuli are grouped under a common response, error reduction is considerably greater than when Stimuli 1 and 2 are combined.

Another factor determining the E_r curves was *O*'s preference for using a single unique response for the stimuli he was most certain of. This preference was revealed by examining the records for the individual *O*s. The preference was most marked when the number of responses exceeded the number of stimuli. This tendency on the part of *O*s is at least in part responsible for E_r being relatively low for the first few stimulus categories when the number of responses exceed the number of stimuli. Its net effect is to make performance under these circumstances more similar to performance when the number of responses and stimuli are equal, insofar as the first few stimuli are concerned.

Effects of increasing the stimulus dimension.—As noted earlier, increasing the range of square sizes from 2–42 to 2–82 mm. sq. resulted in a highly significant improvement in discrimination. However, the magnitude of the increase was not large. The average I_t for the 2–42 mm. sq. range was 1.84 bits and for the 2–82 mm. sq. range 2.01 bits. In terms of the number of categories that could be discriminated without error, this represents an increase of less than half a category. Although the size dimension was doubled in terms of a square root function, it is apparent that nothing approaching a doubling was involved in the subjective scale of size. Pollack (**10**) has reported a similar finding for the pitch dimension. Even when the range of frequencies was increased by a factor of about 20, there was only a slight increase in the information that *O*s could transmit per stimulus presentation.

In Fig. 3 we have scaled the size dimension on an equal discriminability scale after the method described by Garner and Hake (**4**). A separate scaling was performed with the 21 stimulus–21 response and 11 stimulus–11 response data for both size ranges. A scaling based upon the 5 stimulus–5 response data was not possible since there were too few errors. To facilitate comparison, all four curves are plotted so as to have the same E-D value for the 26 mm. sq. stimulus.

The functions obtained with 11 and 21 stimuli in the 2 to 42 mm. sq. range and the 2 to 82 mm. sq. range are quite similar, and differences are so minor as to be attributed to chance errors. The functions obtained for the 2–42 mm. sq. range and the 2–82 mm. sq. range are similar to one another in the low end of the range but they begin to diverge about the 30 mm. This divergence appears due to the upper anchoring effect for the 2–42 range which begins to exert its effect at this point. It is to be noted that the upper anchoring effect is not

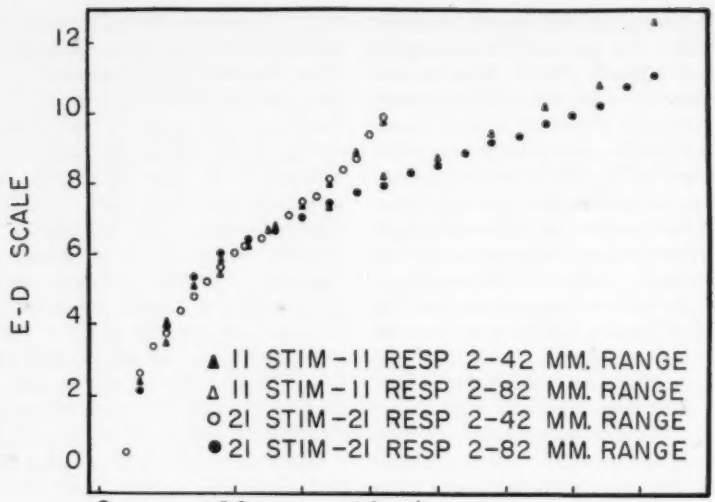

E-D SCALE
SQUARE ROOT OF AREA IN MM
▲ 11 STIM-11 RESP 2-42 MM. RANGE
△ 11 STIM-11 RESP 2-82 MM. RANGE
○ 21 STIM-21 RESP 2-42 MM. RANGE
● 21 STIM-21 RESP 2-82 MM. RANGE

FIG. 3. The equal discriminability scale obtained for 21 stimuli–21 responses and 11 stimuli–11 responses for both the 2–42 and the 2–82 mm. sq. stimulus range. The scaling was based upon the pooled data for the five *Os* in each of these four conditions. Thus the E-D value for each stimulus is based upon responses to 100 presentations of the stimulus. All four curves have been plotted so as to have the same E-D scale value for the 26 mm. sq. stimulus. The 2 mm. sq. stimulus has not been plotted because of the inherent unreliability involved in determining its scale value.

as pronounced for 21 stimuli in the 2–82 mm. sq. range as it is for the other three cases.

The anchor effects suggest why increasing the size range does not produce a larger increment in our measure of discrimination. Since the curves in Fig. 3 are quite similar to those obtained by Garner (2) with loudness judgments, it would appear that the method of absolute judgment is such that a disproportionately large amount of the discrimination that occurs is carried by a relatively few stimuli on each end of the stimulus continuum. If we increase the range of the stimuli, we are in effect adding only middle stimulus categories which are the poorest in conveying information or adding to discrimination. The end or anchor effects may be expected to remain unchanged if the range is increased.

DISCUSSION

Perhaps the most important characteristic of the method of absolute judgment is the prominent reduction in errors or improvement in discrimination that is obtained for the stimuli on the ends and close to the ends of the stimulus series. In the present study we have found that the effects of two major variables, number of response categories and the extent of the stimulus dimension, were markedly influenced by these end or anchor effects. Garner (3) has suggested that these anchor effects may be an artifact of the method of absolute judgment. Response errors to the end stimuli can only be made in one direction. Errors for the lowest stimulus can be made only by calling it too large, and for the large

stimulus only by calling it too small. With a normal distribution of errors, we might theoretically expect errors to be only about half as large for the end stimuli. How far this artificial error reduction might be expected to extend into the middle of the stimulus range would depend upon how great was the dispersion of errors around the correct judgment. The less the error dispersion, the fewer the number of stimuli involved in the anchor effect.

Our data, however, suggest that the anchor effect is more than an artifact. The data presented in Tables 2 and 3 indicate that the judgmental accuracy obtained for the lowest stimulus categories is much greater than would be expected if curtailment of response was the only factor operating. For the first stimulus there are practically no errors, even though the method does not limit errors in the direction of calling this stimulus too large. If curtailment was the only factor involved in producing high judgmental accuracy for this stimulus, we should expect about the same number of "too large" responses for this stimulus as were obtained with the second stimulus category. By the third stimulus category there is opportunity for *O*s to err by calling this stimulus "too small," but as we have seen errors of this kind were quite few in number.

An even more cogent argument is found when we examine the errors and modal responses for the 21–21, 2–42 mm. sq. condition. Here the modal response for the largest stimulus (Stimulus 21) was 15. Since the modal response was so low, there existed little or no artificial curtailment of responses. Yet, as seen in Fig. 3, there is still a prominent anchor effect when the data for this condition are scaled.

While the artificial limiting of responses may be one factor in producing anchor effects, the present data indicate that the end stimuli also serve as subjective standards for the *O*s. In this capacity they would be most effective in improving the judgment or discrimination of the stimulus categories that were closest to them in the stimulus series. This would lead, of course, to the same kind of anchor effects as were obtained here and elsewhere (2).

For the end stimuli to be able to serve as standards for the *O*s, it is necessary that *O* be able to recognize them when they occur. One factor that would determine their ease of recognition would be their distance from the adjacent stimuli on the particular stimulus dimension. The greater the separation, the more frequently *O* will recognize them when they occur and the greater will be the end or anchor effects. Such an explanation would account for the difference in anchor effects obtained between the 11 and 21 stimulus categories for the size range 2–82 mm. sq. With 11 stimulus categories the squares increased by 8 mm. per side whereas with 21 stimuli the increase was only 4 mm. per side. As would be expected, the upper anchor effect was more pronounced for the 11 stimulus condition. If we assume that the difference limen is less for squares in the range of 35 to 40 mm. sq. than it is in the range of 75–80 mm. sq., this explanation can also be extended to cover the finding that the upper anchor effects were greater with the 2–42 mm. sq. range than with the 2–82 mm. sq. range.

Summary

The present experiment was concerned with the effects of three variables upon discrimination and judgment under the absolute method. Discrimination among different sized squares was studied when: (*a*) the size of the squares varied from 2–42 mm. sq. and from 2–82 mm. sq.; (*b*) 5, 11, and 21 different sized squares were presented for judging; and (*c*) 5, 11, and 21 response categories were available for expressing judgments. The data were analyzed along the line of information theory, with discrimination being measured in terms of transmitted information (I_t).

In statistical analysis two significant effects were obtained. The first of these pertained to the stimulus range. Increasing the range from 2–42 to 2–82 mm. sq. resulted in a reliable increase in I_t, but the magnitude of the increase was slight, averaging less than .2 bits.

The second significant effect involved an interaction between the number of stimulus and response categories. It was found that discrimina-

tion as measured by I_t remained constant as the number of stimulus and response categories was increased, provided the number of response categories equalled or exceeded the number of stimulus categories. There was a loss in discrimination when the number of response categories was fewer than the number of stimuli to be judged.

Further analysis of the data suggested (*a*) that both the above effects could be attributed to the end or anchor effects that appear to characterize the absolute method, and (*b*) that the anchor effects obtained with the absolute method cannot be completely accounted for on the basis of the artificial curtailment of responses to the end stimuli.

References

1. Eriksen, C. W. Multidimensional stimulus differences and accuracy of discrimination. *USAF, WADC Tech. Rep.*, 1954, No. 54–165.
2. Garner, W. R. An equal discriminability scale for loudness judgments. *J. exp. Psychol.*, 1952, **43**, 232–238.
3. Garner, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46**, 373–380.
4. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
5. Hake, H. W., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. exp. Psychol.*, 1951, **42**, 358–366.
6. Hunt, W. A., & Volkmann, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49**, 88–92.
7. Johnson, D. M. A systematic treatment of judgment. *Psychol. Bull.*, 1945, **42**, 193–224.
8. McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97–116.
9. Norton, D. W. An empirical investigation of some effects of non-normality and heterogeneity of the *F*-distribution. Unpublished doctor's dissertation, State Univ. of Iowa, 1952.
10. Pollack, I. The information of elementary auditory displays. *J. acoust. Soc. Amer.*, 1952, **24**, 745–749.
11. Volkmann, J. The compression of an absolute scale. *Psychol. Bull.*, 1938, **35**, 676. (Abstract)
12. Volkmann, J., Hunt, W. A., & McGourty, M. Variability of judgment as a function of stimulus-density. *Amer. J. Psychol.*, 1940, **53**, 277–284.
13. Wever, E. G., & Zener, K. E. The method of absolute judgments in psychophysics. *Psychol. Rev.*, 1928, **35**, 466–493.

(Received June 11, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

THE EFFECT OF FAMILIARIZATION UPON SERIAL VERBAL LEARNING[1]

CLYDE E. NOBLE

Louisiana State University

Preceding investigations in this series have been concerned with developing a behavioristic analysis and scale of meaning (**12**), determining the effect of meaning upon serial verbal learning (**13**), measuring familiarity and determining the meaning-familiarity relationship (**14**), and evaluating the dependence of familiarity upon actual frequency of experience (**15**).

It was found that the familiarity (f) and meaning (m) scales had high reliability and were closely correlated (**12, 14**). Experiment then revealed that serial verbal learning was significantly facilitated when the materials were composed of items high in m value (**13**). It follows that this type of performance is also a positive function of f. Now, the f value of verbal concepts has recently been shown (**15**) to be determined almost uniquely by frequency of exposure and articulation (n). On a purely empirical basis, therefore, mere familiarization with the individual items of a list should enhance proficiency in later serial learning. A test of this hypothesis, sometimes known as Robinson's "law of acquaintance" (**19**), is the object of the present experiment.

Two difficulties are encountered in formulating an exact deduction from available data. First, as the writer has emphasized previously, the status of the familiarity concept in his original scaling investigation (**14**) was that of an *individual-difference* variable, whereas the f values reported in the most recent study (**15**) were *experimentally* produced by deliberate manipulation in the laboratory. Let us denote these concepts by the symbols f_i and f_e, respectively. The problem which thus arises is whether one can justifiably coordinate these two distinct definitions of familiarity. Second, the earlier discussion of the f scale (**14**) called attention to its interval properties, whereby mathematical substitution or other precise deductive techniques are subject to limitations inherent in the scaling technique. This leads to certain indeterminacies in the derivation of exact functional relationships involving the response variable f.

Concerning the first question, in view of the identity of the defining operations and of the population from which the samples of judges were drawn in these studies, we shall tentatively assume the proportionality of f_i and f_e. With respect to the second problem, it is predicted that the rate of serial learning will be some positive function of n. Curvilinearity and negative acceleration are expected characteristics of the facilitation function, but its general form and specific parameters remain an empirical matter.

METHOD

General.—The plan of this experiment was to present visually to S a set of relatively unfamiliar and meaningless verbal stimuli in vary-

[1] The experimental work for this study was performed as part of the program of the Skill Components Research Laboratory, Air Force Personnel and Training Research Center. S/Sgt. R. B. Card supervised its administration.

ing frequencies of occurrence and in a random sequence. Upon each presentation *S* was required merely to pronounce the item aloud, thus insuring stimulus reception as well as exercise of the response. Two minutes later, six of these items were placed in serial lists to be learned in counterbalanced order by the anticipation method. By comparing the performance of groups differing only in their prelearning familiarity with the material, other things constant, the effect of frequency of experience upon verbal learning may be determined.

Apparatus.—The exposure device consisted of a 48-item 2 × 2-in. automatic slide projector (Selectroslide). Stimulus duration, controlled by Hunter-Brown electronic timers, averaged 3 sec. with a 1.44-sec. dark period between stimuli. This gave a presentation rate of one slide every 4.44 sec. for all conditions. Fifteen items were chosen from the low extremes of Noble's *m* and *f* scales (**12, 14**). The following eight frequencies (*n*) were then assigned to each of six of these words selected at random: 0, 1, 2, 3, 4, 5, 10, 20. Initially these stimuli had average *m*- and *f*-scale values of 1.27 and .44, respectively. The remaining nine items were used as fillers in order to equalize the total number of exposures for each *S*. All words were reproduced in block letters and prepared as slides.

For the *familiarization* phase of the experiment, three cylindrical projector cartridges were employed. The six experimental stimuli were mounted at random serial positions, each cartridge containing exactly one-third of the exposures required for a given treatment condition. Intervening spaces were filled with the nonexperimental items, repeated as necessary. Thus each *S* had 48 × 3 = 144 total exposures. This precaution served as a control for differential transfer effects due to general task factors, such as learning-to-learn, set, or warm-up (**9**). For the *learning* phase, a fourth cartridge was used. It contained the six experimental items together with a cue symbol and a blank slide to constitute a trial sequence. The intertrial interval was 8.88 sec.

Procedure.—The experimental plan utilized an 8 × 6 factorial design. Five replications were run, in which the interaction of treatment (*n* value) and order (list sequence) effects was confounded with individual differences. Each of the eight familiarization conditions was followed by six-link serial verbal learning by the anticipation method. The lists were presented in all six orders of one random permutation. This insured that any differential effects of the various sequences of exposure during familiarization would be largely neutralized by the counterbalancing procedure of the learning phase. Some may regard this as a control for so-called "incidental" learning, the relevant phenomenon being that of remote associations (**9**). Although the present design does not provide for complete counterbalancing, it is considered adequate for the purpose.

The *S*s consisted of 240 basic airmen at Lackland Air Force Base, assigned without bias to the eight treatments within each of the six order conditions. Five *S*s thus appeared in each treatment-order combination, giving 30 cases per treatment. For the familiarization phase of the experiment, *S* was required simply to pronounce each word, as in the previous study (**15**). For the learning phase, the instructions were those of the typical serial anticipation procedure, with correction for errors. Anticipations began on Trial 1, the second repetition of the list. On the first presentation *S* pronounced only. The mastery criterion was set at two successive error-free trials.

Rest periods of exactly 2 min. were given after each group of 48 exposures during the familiarization phase, and also prior to the instructions for the learning phase. The *S*s were tested individually during both parts of the experiment and were not informed of its purpose. They were also isolated from the remaining pool of *S*s in order to prevent discussion of the procedure.

Results

The general results are summarized in Fig. 1. Successive correct anticipations are plotted as functions of mean trials for the eight frequency conditions. It is evident that *n* has a facilitative effect upon serial verbal learning in accordance with prediction. The acquisition curves for the 0, 10, and 20 conditions are most clearly separated, with those for 1 through 5 interlacing between 0 and 10.

A convenient significance test of the differences among the mean trials to mastery is provided by Friedman's (**3**) nonparametric analysis of variance. This test was performed on the 8 × 6 matrix of criterion scores, in which each cell entry was a rank based on a mean of 5 measures. For $df = 7$, $\chi^2 = 172.4$, which is highly significant ($p < .001$).

Since differential effects of fami-

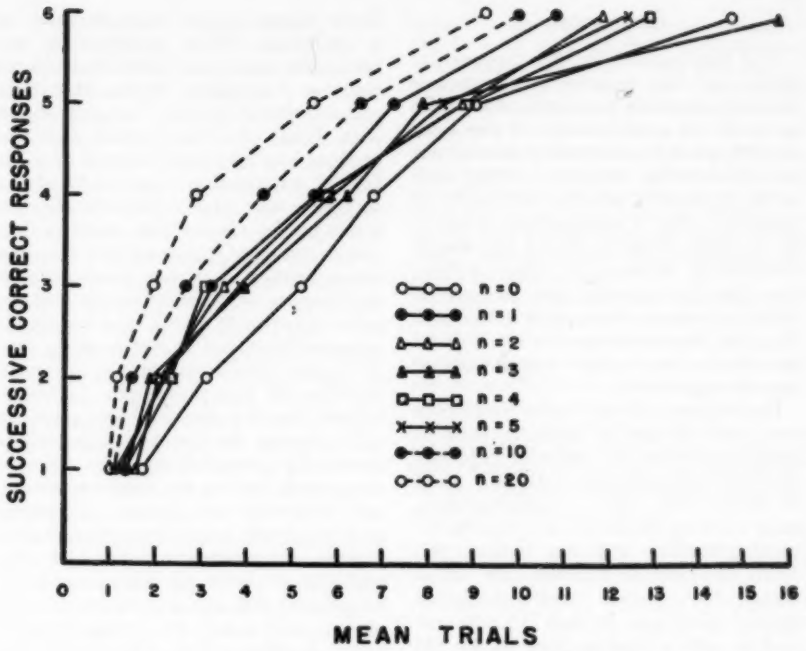


FIG. 1. Successive correct responses as functions of mean trials for eight conditions of prior frequency of experience (n).

liarization upon performance should be most clearly revealed at the outset of learning, it is instructive to examine the data for Trial 1 in more detail. A sensitive response measure is given by the percentage of correct anticipations for the eight groups. Figure 2 indicates this facilitation function for the *Ss*' initial encounter with the task. The significance of the observed trend is evaluated by simply testing the difference between the two extreme percentages, 10.6% and 44.8%. The difference is 34.2%. By a z test for $df = 58$, the trend is shown to be highly significant ($p < .001$).

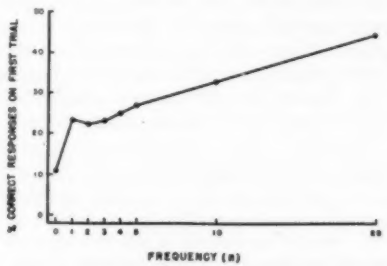


FIG. 2. Percentage of correct responses on Trial 1 as a function of prior frequency of experience (n).

One should recall at this point that no anticipations were required on the first presentation of the list. By Trial 1, however, all *Ss* had received one reinforcement for each of the six items. For this reason the percentages in Fig. 2 have presumably shifted upward by an unknown amount. The shift continued as practice increased, with a slope approaching zero.

Discussion

The function in Fig. 2 is apparently curvilinear, but negative acceleration is not evident within the range of n values sampled. A consideration of the limit of 100% correct responses lends credence to a diminishing returns principle, such as the hyperbola relating familiarity to frequency (15). Careful graphical analysis, however, failed to reveal any simple relationship between f values derived from the latter study and the probabilities of correct choice shown in Fig. 2. This fact does not support a mediational hypothesis, but other interpretations may be considered.

Hypotheses of one type emphasize perceptual factors in learning, such as predifferentiation (4), selective attention (8), cue distinctiveness (11), and speed of recognition (21). Unfortunately, many learning studies are complicated by learning-to-learn and set factors (9). When appropriate controls are introduced, the alleged predifferentiation transfer effect may be zero (1). Recent work in serial verbal learning (6, 13, 18) has led to hypotheses of a second type, suggesting that facilitation is somehow associated with responding. One possibility is that the familiarization procedure imposes restrictions upon response availability (16), such that each S's hierarchy of verbal associations becomes ordered with respect to the ratio $n/144$. If, by generalization, the more strongly conditioned responses have initially a greater probability of being elicited by any intralist stimulus, the degree of facilitation would increase with amount of prior experience. This is also consistent with related empirical data (10, 21, 23) connecting reaction latency to relative frequencies of S or R. Hovland and Kurtz (6) have advanced a similar proposal by analogy to Shannon's (20) mathematical theory of communication.

Consider last the hypothesis that the present results are due to differential practice in pronunciation. With improvement in the motor skill of articulation, fewer anticipations would be mispronounced on Trial 1 as well as thereafter. Error scores would then decrease as n increased. Prior practice in verbalization might also serve to attenuate response variability during the initial presentation of the list. Let us postulate with Hull (7) that verbal responses produce characteristic internal cues (s). In the anticipation learning situation, such cues may interact with the stimulus traces left by the intralist items to form compound conditioned stimuli ($\bar{s}$). Assume further that the proprioceptive components of these stimulus patterns grow increasingly stable with training in pronunciation, but at a decreasing rate. Now, from Hull's postulate of the generalization of habit strength ($_sH_R$), it follows that the effective habit strengths ($_s\bar{H}_R$) during the learning phase will be maximally controlled by those stimulus compounds having the least variability, and minimally conversely. Expressed in more simple terms, strength of conditioning is dependent upon stimulus constancy.[2] From the foregoing, it can be deduced that initial proficiency in the present task should be a positive curvilinear function of n. One could also invoke this principle to explain the writer's (13) demonstration of the facilitative role of meaning (m) in serial verbal learning. The fundamental fact seems to be that words of higher m value are more uniformly pronounced. It also happens that the difficulty-meaning relationship, to a first approximation, is a decreasing curvilinear function. This is consistent with theoretical expectation.

A resolution is now possible of the apparent contradictions among previous experimental tests of Robinson's so-called "law of acquaintance" (19). With respect to verbal retention, Peters (17) found no differences employing Thorndike's 1921 word count. Using

[2] The phenomenon of external inhibition in Pavlovian conditioning experiments provides a classical illustration of the decremental effects upon behavior of variability in a stimulus complex. Hull (7) conceived of this as one of the external sources of behavioral oscillation ($_sO_R$), which theoretically subtracts from effective reaction potential ($_s\bar{E}_R$) to cause a reduction in performance.

the 1944 Thorndike-Lorge count, however, Hall (5) extended Peters' range of values and obtained significant facilitation. When familiarization is induced experimentally, as Bousfield, Sedgewick, and Cohen (2) have done in another connection, the outcome is similar. In serial verbal learning, using frequencies of 0 and 3, Waters (22) was unable to demonstrate a difference uncomplicated by transfer effects arising from the orders of item presentation in the familiarization and learning phases. He concluded, and McGeoch and Irion (9) agreed, that Robinson's principle was not verified. Hovland and Kurtz (6) succeeded in controlling the sequential factor, but their positive results were obtained with frequencies of 0 and 30 (or greater). This left unexplored the nature of the functional relationship. An experiment by Price and Lewis (18) is relevant to the pronunciation hypothesis, but it was not a parametric investigation.

Inspection of Fig. 1 and 2 suggests that the data of all these studies may actually lie on a continuum, and that they agree with the hypothesis that proficiency in verbal chaining is positively related to *n*. This proposition, it should be noted, is not the inductive generalization advanced (on very limited evidence) by Robinson. Rather it is the specific deductive consequence of a series of related experiments (12, 13, 14, 15) designed to isolate the S-R components in a set of R-R laws pertaining to verbal behavior. The proposition is also derivable independently from a general S-R behavior theory, as outlined above. Actually, Robinson's principle was never expressed in clear enough terms to permit an unequivocal test. The reason is that "acquaintance" consists of two logically separable components: familiarity and frequency. The one is psychological, the other physical (15).

Summary

From previous research it was deduced that performance in serial verbal learning should be some positive, curvilinear function of frequency of experience with the items to be learned. This hypothesis was tested by presenting visually to 240 *Ss* a set of six relatively unfamiliar and meaningless verbal stimuli in the following frequencies: 0, 1, 2, 3, 4, 5, 10, 20. Each item was pronounced as it occurred, thus insuring stimulus reception and exercise of the response. By appropriate use of filler words, all *Ss* had the same total number of exposures. Two minutes later the six experimental items were encountered in serial lists to be learned by the anticipation method to a criterion of two successive correct trials. Each of the eight independent treatment groups contained 30 *Ss*.

The effect of familiarization was a significant reduction in the number of trials required for mastery. A similar finding was true of the percentage of correct responses on the initial trial, the degree of facilitation increasing curvilinearly with amount of prior experience.

The results were discussed in terms of a reinterpretation of Robinson's "law of acquaintance." Several explanations were considered, the most promising being concerned with differential practice in pronunciation.

References

1. Arnoult, M. D. Transfer of predifferentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, **45**, 401–409.
2. Bousfield, W. A., Sedgewick, C. H. W., & Cohen, B. H. Certain temporal characteristics of the recall of verbal associates. *Amer. J. Psychol.*, 1954, **67**, 111–118.
3. Friedman, M. Use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
4. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
5. Hall, J. F. Learning as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.
6. Hovland, C. I., & Kurtz, K. Experimental studies in rote-learning theory: X. Prelearning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.
7. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
8. Kurtz, K., & Hovland, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157–164.

9. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (Rev. Ed.) New York: Longmans, Green, 1952.
10. Merkel, J. Die zeitlichen Verhältnisse der Willensthätigkeit. *Phil. Stud.*, 1885, **2**, 73–127.
11. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
12. Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.
13. Noble, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446; 1952, **44**, 465.
14. Noble, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.
15. Noble, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, **47**, 13–16.
16. Noble, C. E. Compound trial-and-error learning as a function of response availability (N_R). *J. exp. Psychol.*, 1955, **49**, 93–96.
17. Peters, H. N. The relationship between familiarity of words and their memory value. *Amer. J. Psychol.*, 1936, **48**, 572–584.
18. Price, H. G., & Lewis, D. Increased pronouncing behavior as a factor in serial learning. *J. exp. Psychol.*, 1954, **47**, 95–100.
19. Robinson, E. S. *Association theory today.* New York: Century, 1932.
20. Shannon, C. E. A mathematical theory of communication. *Bell Syst. tech. J.*, 1948, **27**, 379–423, 623–656.
21. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.
22. Waters, R. H. The law of acquaintance. *J. exp. Psychol.*, 1939, **24**, 180–191.
23. Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* (Rev. Ed.) New York: Holt, 1954.

(Received July 12, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

EFFECTS OF THREE TYPES OF PRETRAINING ON DISCRIMINATION LEARNING IN PRESCHOOL CHILDREN [1]

GORDON N. CANTOR

Child Welfare Research Station, State University of Iowa[2]

The prediction that the possession of names for the stimuli in a learning task will enhance performance on that task has been tested in several studies. The enhancement phenomenon has been referred to as the "acquired distinctiveness of cues" (**9, 10**) and "stimulus predifferentiation" (**2, 3, 5**). In general, the experimental paradigm has involved a transfer task situation in which *Ss* are required to respond to certain discriminable stimuli. This task is preceded by a pretraining situation in which some of the *Ss* are taught names for these same stimuli, the names themselves having no connotations with respect to the nature of the responses in the transfer task. This type of pretraining has been termed "relevant-S." The more recent studies have utilized "irrelevant" pretraining groups, as well as or in place of "no-pretraining" groups, in order to control for warm-up and learning-to-learn effects (**2, 3, 6, 8**). The need for such controls has been demonstrated (**4**). Performances of relevant-S groups on various types of transfer tasks have been shown to be superior to those of irrelevant groups (**2, 3, 6**).[3]

The present experiment is concerned with the relative performances on a transfer task of three groups of children. Two of these involve the conventional relevant-S and irrelevant varieties of pretraining. The third—to be called hereafter the "attention" group—is made to attend to the relevant transfer task stimuli during a pretraining session, but is not given any verbal learning.

It is predicted that the relevant-S group will surpass the irrelevant and attention groups in performance on the transfer task. No prediction is made with respect to the relative performances of the irrelevant and attention groups.

METHOD

Subjects.—The *Ss* were 60 children from the State University of Iowa Preschools and the Iowa City Parents' Preschool. There were 10 3-yr.-olds, 38 4-yr.-olds, and 12 5-yr.-olds. An upper and a lower age group were constituted and *Ss* from within each age level were randomly assigned to the three treatment groups.

Procedure.—The materials used in the verbal pretraining consisted of photographs of pen and ink sketches of the following: (*a*) a pair of human adult female faces; and (*b*) a pair of human juvenile male faces.

The *Ss* in the relevant-S group received relevant-S pretraining in which they learned to associate the names "Jean" and "Peg" with the female faces (Pair A); learning was continued until a criterion of 15 consecutive pairs of correct responses was met. The *Ss* in the irrelevant group were given irrelevant pretraining, involving the male faces (Pair B) and the names "Jack" and "Pete," to the same criterion. The stimuli in these verbal learning situations were presented side by side to *S*. The right-left relationships of the pictures were varied in a prearranged and nonalternating fashion. An imi-

[1] This paper is based on a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is indebted to Dr. Charles C. Spiker for his generous guidance.

[2] Now at George Peabody College for Teachers.

[3] The terms "relevant-S" and "irrelevant" are McAllister's (**8**). The latter term refers to pretraining in which the stimuli are different from those occurring in the transfer task.

tation coin reward was presented to *S* for each correct pair of responses. The *S* was corrected when erroneous responses were given. When the criterion was met, each *S* was allowed to cash in the accumulated coins for one-half of a toy set.

The attention pretraining did not involve verbal learning; rather, the *S*s in the attention group were made to point to various parts of the relevant stimuli (Pair A). The stimuli were presented in a manner comparable to that involved in the relevant-S and irrelevant pretraining. Thirty trials were given to each attention group *S*, however, and a coin reward was given each time *S* pointed to the designated parts of the pictures in what *E* judged to be a "careful" manner. If *S* appeared lax in his pointing responses, the coin reward was withheld and *S* was urged to exercise more care.

All *S*s were treated identically in the transfer task, which was administered immediately following pretraining. The apparatus and procedures of a recently reported study (1) on discrimination learning in preschool children were utilized, with slight modifications, in the transfer task. The stimuli were two wooden cars, identical in appearance except that each had a different one of the two faces from Pair A mounted on its side. The *S*'s task involved choosing one of the two cars and rolling it down a track. The *S* was led to believe that the car's striking a dummy switch at the end of the track caused a marble dispenser to be activated. Actually, *E* controlled the operation of the dispenser by means of a hidden button. For each *S*, one of the two cars was arbitrarily designated by *E* as the "correct" stimulus. When this car was chosen by *S* and run down the track, a marble reward was dispensed; when the "incorrect" car was selected, no marble reward was delivered. The choice by *S* of one of the stimuli, followed by the car's run down the track and the subsequent reinforcement or nonreinforcement, constituted a trial. Thirty transfer task trials were given, *S* being instructed to obtain as many marbles as possible. Following Trial 30, *S* was allowed to exchange the accumulated marbles for the remainder of the toy set and the experiment was terminated. The response measure was the number of correct choices made in the transfer task.

Results and Discussion

A treatments-by-levels analysis (7) of the pretraining data revealed that the relevant-S and irrelevant groups did not differ significantly from each other in mean number of trials required to meet the criterion. Table 1 shows the means and *SD*'s for the subgroups. Table 2 presents the over-all means and *SD*'s for the transfer task data.

In order to allow for a treatments × ages × trials analysis, the 30 trials were divided into six blocks of five trials each. A Type III analysis (7)

TABLE 1

Pretraining Results—Number of Trials to Criterion

Age	Relevant-S		Irrelevant	
	Mean	*SD*	Mean	*SD*
L	26.0	10.80	27.3	10.24
U	28.8	14.32	21.8	6.79

TABLE 2

Number of Correct Responses in 30 Transfer Trials

Age	Relevant-S		Irrelevant		Attention	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
L	23.6	5.97	16.5	5.50	18.0	5.44
U	20.9	7.16	17.2	5.44	17.4	3.98

TABLE 3

Summary of Analysis Applied to Treatments × Ages × Trials Data †

Source	*df*	MS	*F*
Between *S*s	59		
Treatments	2	28.10	4.74*
Ages	1	1.88	<1.00
Error (*b*)	54	5.93‡	
Within *S*s	300		
Trials	5	4.25	4.01**
Treatments × Trials	10	2.47	2.33*
Error (*w*)	270	1.06‡	
Total	359		

* $p < .05$.
** $p < .01$.
† All interactions involving age were nonsignificant.
‡ On the assumption of homogeneity of variance in the cells, the error terms obtained in this analysis, together with their respective *df*, were utilized in the analyses to be reported subsequently.

was applied to these data, the results of which are summarized in Table 3.

As indicated in Table 3, the treatments effect was significant at the .05 level of confidence, as was the treatments × trials interaction. Since no interaction involving age was significant, it was possible to compare the three treatment groups, taken two at a time, disregarding age levels. This was accomplished by the use of three Type I designs (7), the results of which are summarized in Table 4.

These analyses indicate that the relevant-S group surpassed both the irrelevant and attention groups in mean number of correct choices. The former comparison yielded a difference significant at the .01 level, whereas the latter difference was significant at the .05 level. The irrelevant and attention groups did not differ significantly from each other in terms of the treatment effects.

The treatments × trials interaction was significant at the .01 level for the relevant-S vs. irrelevant data and at the .05 level for the irrelevant vs. attention data, indicating that the learning curves differed from each other for the treatment groups involved.

TABLE 4

SUMMARY OF ANALYSES APPLIED TO THE DATA COMPARING THE TREATMENTS TWO AT A TIME, DISREGARDING AGE LEVELS

Source	*df*	Mean Squares		
		Relevant-S vs. Irrelevant	Relevant-S vs. Attention	Irrelevant vs. Attention
Between *Ss*				
Treatments	1	48.60**	34.51*	1.21
Error (*b*)	54	5.93	5.93	5.93
Within *Ss*				
Trials	5	4.53**	4.43**	2.01
Treatments × Trials	5	3.65**	.81	2.95*
Error (*w*)	270	1.06	1.06	1.06

* $p < .05$.
** $p < .01$.

The results indicate that the experience of learning names for a pair of stimuli evidently did facilitate later performance on a task involving those stimuli, as predicted. By the use made of the irrelevant and attention groups, it was shown that the enhancement in performance attributable to the possession of names was greater than that which could be attributed either to the transfer of nonspecific reactions or to mere exposure to the relevant stimuli during pretraining.

SUMMARY

Three groups, each consisting of 20 3- to 5-yr.-old children, were given identical treatment in a transfer task situation after having experienced different types of pretraining. The relevant-S group learned during pretraining to associate names with each of the two pictures of faces that subsequently served as stimuli in the transfer task. The irrelevant group was given a comparable amount of paired-associate learning, but with stimuli unrelated to the transfer task. The attention group was not given a verbal learning experience, but rather was merely exposed to the relevant stimuli during pretraining. The transfer task involved a simple discrimination learning situation, the response measure being the number of correct choices made in 30 presentations of the relevant stimuli.

Analyses of the transfer task data revealed that the relevant-S group performed significantly better than did the irrelevant and attention groups. Significant treatments × trials interactions indicated that the learning curves of the relevant-S and irrelevant groups differed, as did the curves for the irrelevant and attention groups.

It was concluded that the results were in agreement with the prediction that the possession of names for the stimuli in a learning task would enhance performance on that task.

REFERENCES

1. CANTOR, G. N., & SPIKER, C. C. Effects of nonreinforced trials on discrimination learning in preschool children. *J. exp. Psychol.*, 1954, **47**, 256–258.
2. CANTOR, J. H. Amount of pretraining as a factor in stimulus pre-differentiation.

Unpublished doctor's dissertation, State Univ. of Iowa, 1954.

3. Dysinger, D. W. An investigation of stimulus pre-differentiation in a choice discrimination problem. Unpublished doctor's dissertation, State Univ. of Iowa, 1951.
4. Farber, I. E., & Murfin, F. L. Performance set as a factor in transfer of training. Paper read at Midwest. Psychol. Ass., Chicago, April, 1951.
5. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
6. Gerjuoy, I. R. Discrimination learning as a function of the similarity of the stimulus names. Unpublished doctor's dissertation, State Univ. of Iowa, 1953.
7. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
8. McAllister, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.
9. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
10. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.

(Received July 16, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

FACTORS INFLUENCING RATE AND EXTENT OF LEARNING IN THE PRESENCE OF MISINFORMATIVE FEEDBACK [1]

ROBERT E. MORIN [2]

University of Wisconsin

Many responses have to be learned in the face of inconsistent consequences. Even the most adequate response to a particular stimulus may not always be followed by reward. Misinformative feedback is said to occur when an optimum response, one which provides the greatest expected return over a series of trials, is followed by a consequence or signal which informs the responder he has made an incorrect response, or when a nonoptimum response results in a signal that the response was correct. The experiment reported here was designed to determine how acquisition in a double-risk or two-choice response situation depends on percentage of misinformative feedback, awareness on the part of *Ss* of possible misinformative feedback, the presence of a correction signal in the feedback channel on misinformative trials, and an explanation of the function of the correction signal.

There are many similarities between the task used in this study and that of the Humphreys-type verbal conditioning or guessing game experiments (**4, 6, 7, 8, 10**), Brunswik's work on partial reinforcement with rats (**2**), and studies of gambling and problem solving by Jarrett (**9**). Typically *S* makes one of two possible responses, only one being correct on each trial, but neither being correct on all trials. A left turn in a T maze may be rewarded 75% of the time, a right turn 25% of the time, with the sequence of left and right rewards randomly determined. Or *Ss* may try to predict which of two lights will light when in fact the left light comes on 75% of the trials and the right light 25% of the trials. The optimum behavior in such circumstances, the behavior which maximizes the proportion of successful trials, is a left response on 100% of the trials. There is misinformative feedback in the system on the 25% of the trials for which a choice of *left* is not rewarded or a choice of *right* is rewarded. The most frequent finding in these studies is that *Ss* do not reach optimum behavior but instead level off at a probability matching solution (**4, 6, 7, 9**). That is, the responses are divided among the choices such that a response is made about as often as it is rewarded. In terms of the above examples, the proportion of right choices approximates 25% and the proportion of left choices approximates 75%.

Work by Jarrett (**9**) suggests that *S*'s conceptualization of the task may be an important variable in determining whether probability matching

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at the University of Wisconsin. The writer is greatly indebted to Drs. David A. Grant, thesis director, Willard Thurlow, and L. Benjamin Wyckoff. This research was supported in part by the United States Air Force under Contract No. AF 33(038)-23294 monitored by the Armament Systems Personnel Research Laboratory, Air Force Personnel and Training Research Center, Lowry Air Force Base, Denver, Colorado. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at the University of Texas.

solutions are surpassed and optimum solutions are reached. She found that more *Ss* behaved optimally at a gambling task which paid off money than at a task structured as a learning problem and using complex visual forms as stimuli. In the two tasks *Ss* had different notions about the possibility of a uniformly successful solution, but there were also many other differences between the tasks. Still the idea that *S*'s conceptualization of the task may be important to acquisition is an appealing one and is investigated further in this study by comparing performances under conditions of awareness and nonawareness of possible misinformative feedback. It seems reasonable that *Ss* aware of misinformative feedback will more often accept solutions which do not provide success on every trial. If this is the case, ultimately more *Ss* should arrive at optimal solutions, but errors prior to solution should be more persistent and systematic since incorrect hypotheses should be perpetuated longer. Strategies which lead to temporary chance success or partial but not optimum success should be rejected less easily.

The presence of a correction signal in the feedback channel on misinformative trials should further support optimum behavior by providing an element of consistency. It could lead to a redefinition of success by some *Ss* if they correctly surmised its function. The behavior of *Ss* given an unexplained correction signal is compared with that of *Ss* who know the function of the signal.

Method

Experimental conditions.—Twenty-four groups were run in a 6 × 4 orthogonal design. They were given 0, 10, 20, 30, 40, or 50% misinformative feedback. A second categorization depended on instructions and the presence or absence of the correction signal. The four levels of this second variable are designated B, BA, BAC, and BACE where B as a part of the designation means these groups were given a set of *basic* learning instructions, A indicates *awareness* by *Ss* of possible misinformative feedback, C denotes the presence of an unexplained *correction* signal on misinformative trials, and E refers to the fact that an *explanation* of the function of the correction signal was given.

Apparatus.—The stimuli were 1-in. opal jeweled lamps mounted in a circle of 5-in. radius on a black crackle finished panel. One light was mounted at 0° and the others were placed at 45° intervals. The *S* sat about 12 ft. from the stimulus panel at a bench-table divided into four sections for multiple testing. Each section was provided with a response unit and a feedback display which could be seen only by *S* at that section. The response unit consisted of two keys, one mounted about 3 in. to the left of the other. The red and green feedback lights were 1-in. jeweled lamps placed at the same height and 2½ in. apart on the upper half of the feedback display box which was to the left of the response keys and against the back edge of the table. Two sections had the red light to the left of the green and two had the red light on the right. The lower half of the feedback display box was a 3- × 4-in. white flash glass behind which was a 25-w. incandescent bulb. This was used to provide a correction signal.

The rest of the apparatus was on *E*'s table behind *Ss*' bench-table. The control unit consisted of two Western Union Teletype Transmitters, associated relays, and a synchronous motor mechanical timer. It automatically determined the order and durations of the stimulus events including rest intervals and setting up of the feedback channels. An audio oscillator set at 800 cps signalled *Ss* when to respond and an Esterline-Angus operations recorder gave a complete graphic record of all stimulus events and responses. A tape recorder was used to administer the instructions.

Subjects.—The *Ss* were 288 University of Wisconsin male students. Of these, 192 were selected from a group of students who volunteered to serve as paid *Ss* in experiments at the Laboratory of Experimental Psychology. The other 96 were volunteers from psychology classes who worked for experimental points to be credited to their grades. Though many of the *Ss* had previous experience in psychological experiments, none had served in any related experiments. Of the 12 *Ss* in each condition, 8 were paid and 4 worked for experimental points. The *Ss* were run in groups of four whenever possible, but at times scheduling difficulties made it necessary to run one, two, or three *Ss*.

Procedure.—Tape-recorded instructions were administered to *S*s in a group at the beginning of the experiment. The instructions are paraphrased below to reduce somewhat the redundancies in the instructions as they were presented. All groups (B, BA, BAC, BACE) received the following basic instructions:

"We are interested in determining how quickly you can learn when it is correct to press this left key and when it is correct to press this right key. A trial will begin when one of the white lights at the front of the room comes on. Three seconds later a tone will sound to tell you to select one of the two keys and depress it. If you depress the correct key the green light on this box will light and if you depress the wrong key the red light will light. On every trial one key will be right and one key will be wrong, and on every trial you should depress only one key. You should leave the key down only long enough to notice whether the green or the red light comes on. It is necessary to pay attention to which white stimulus light comes on. You cannot solve the problem merely by repeating a sequence of responses like left, left, right, right, left, left, right, right. After a series of trials you will be given a rest period."

The recorder was shut off at this point for the B groups, but the following was added for the others (BA, BAC, BACE) to make them aware of the possibility of misinformative feedback.

"It may help you to know that we have made the equipment so that it does not always tell the truth. Let us suppose you have pressed the correct key. Then the green light *should* light. But occasionally the equipment may lie to you and a red light will come on even though you were really correct and you deserved the green light. If you press the wrong key you deserve a red light, but the equipment may lie to you so that a green light will come on to make it appear you were correct."

The BAC groups were given an unexplained correction signal on misinformative trials but no additional instructions. Only the BACE groups were given both a correction signal and an explanation of its function. Their instructions continued:

"There is a light behind the white flash glass which will come on near the end of every trial the equipment lies to you."

As soon as a group's recorded instructions were completed, *E* read this addition:

"If all of you solve the problem before the period is up, then all of you may leave early. If any one of you fails to solve the problem, then none of you may leave early. Are there any questions about any of the instructions?"

There were a few questions and these were generally easily answered by paraphrasing the instructions. When someone asked what principles to use he was told it was his job to discover them.

Every trial lasted 9 sec. and intertrial intervals were 1 sec. When a correction signal was necessary it came on 2 sec. after the termination of the 1-sec., 800-cps tone telling *S* to respond, and 6 sec. after the start of the trial. It went off 3 sec. later when the trial ended. The session lasted 240 trials or until all *S*s tested at one time learned to a criterion of 40 consecutive errorless responses, five to each light. Two-minute rests were given after Trials 80 and 160.

Over the series of trials each stimulus light was presented for the nth time before any stimulus was given $(n + 1)$ times and no stimulus was ever its own immediate successor. "Percentage of misinformative feedback" was descriptive for every 10 trials and for individual lights over 80 trials. Sequences of lights and misinformation trials were random within these restrictions.

If the stimulus lights are lettered A to H in a clockwise fashion giving the topmost light the designation A, then lights B, C, F, and G formed one set and lights H, A, D, and E made up a second set. For half the *S*s in each condition left responses were correct for one set and right responses were correct for the other set. For the other half of the *S*s the solution was reversed

Results

All responses were evaluated for correctness against an optimal solution criterion. This meant always responding with the left key to four of the stimulus lights and with the right key to the other four. Even the B and BA 50% misinformative feedback groups for whom the problem was insoluble were judged by this standard. The basic datum and entry in most of the analyses was a count of the number of correct responses over a series of 40 trials. Acquisition curves based on these values combine the acquisition curves for individual stimulus lights. It seemed a reasonable assumption that the same processes should operate for all lights. Figure 1 presents the 24 acquisition curves each based on 12 *S*s. Groups with the same percentage of misin-

formative feedback form subplots and the parametric designations within subplots are B, BA, BAC, and BACE. Mean percentages of correct response appear as a function of successive 40-trial blocks of practice. Details of the curves will be considered as analyses of them are described.

Among the sets of groups designated B, BA, BAC, and BACE each set differs from its immediate predecessor by the cumulative addition of one new element to the experimental situation. These elements are awareness of misinformative feedback, a correction signal on "lie" trials, and an explanation of this correction signal. A separate analysis was done to test the effects of each of these elements and of interactions with percentage of mis-

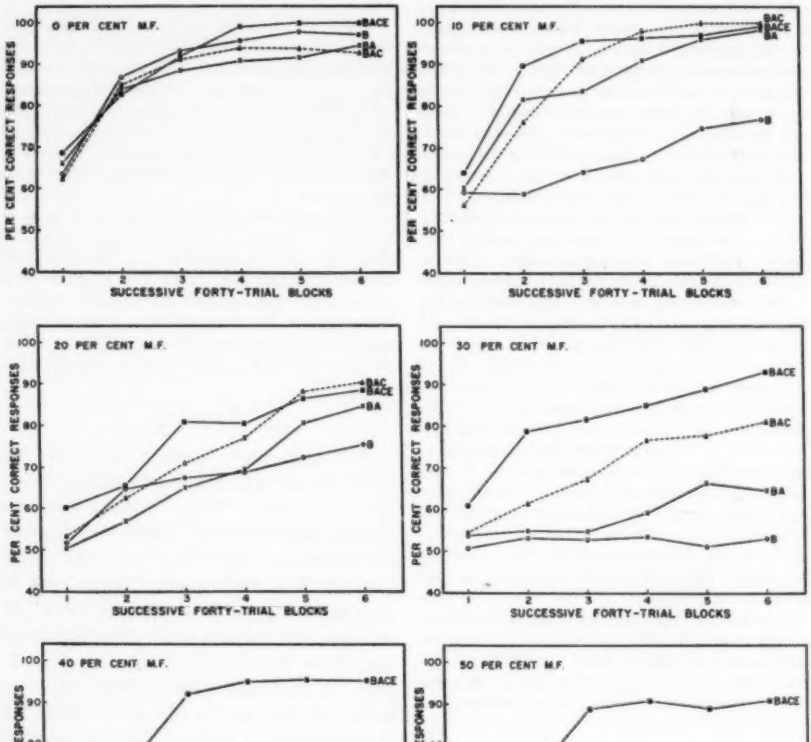


FIG 1. Acquisition curves for the 24 groups presented in subplots for each percentage of misinformative feedback.

TABLE 1

TREND ANALYSES APPLIED TO DATA OF B AND BA GROUPS, BA AND BAC GROUPS, AND BAC AND BACE GROUPS

Source of Variation	*df*	*F* Ratios		
		B-BA	BA-BAC	BAC-BACE
Between group means				
a. Per cent MF (P)	5	53.79**	34.04**	8.72**
b. Column variable (V)†	1	6.01*	6.38*	28.53**
c. P × V	5	3.95**	.58	4.14**
Between group slopes				
d. P	5	17.92**	12.36**	1.52
e. V	1	5.21*	7.28**	1.16
f. P × V	5	2.04	.45	2.95*
Between group quadratics				
g. P	5	4.50**	8.26**	4.12**
h. V	1	3.52	4.12*	7.50**
i. P × V	5	1.20	.53	3.49**
j. Between individual means (error for *a*, *b*, *c*)	132	33.21** (98.05)‡	52.04** (146.32)	75.08** (192.44)
k. Between individual slopes (error for *d*, *e*, *f*)	132	9.14** (26.98)	14.67** (41.26)	18.10** (46.40)
l. Between individual quadratics (error for *g*, *h*, *i*)	132	6.57** (19.40)	6.86** (19.28)	7.86** (20.16)
m. Over-all slope	1	1249.25**	2875.46**	5124.36**
n. Over-all quadratic	1	727.85**	181.70**	748.76**
o. Over-all deviations from linear and quadratic estimation	3	8.00**	7.13**	13.29**
p. Group deviations from estimation	33	3.80**	3.95**	5.68**
q. Individual deviations from estimation (error for *j* through *p*)	1260	(2.95)	(2.91)	(2.56)

* Significant at .05 level of confidence.
** Significant at .01 level of confidence.
† V represents the variable indicated at the top of each column of *F*'s, e.g., in the B-BA column this is awareness vs. nonawareness of misinformative feedback.
‡ Values in parentheses are mean squares.

informative feedback. So one analysis compared the six B groups with the six BA groups; a second compared BA and BAC groups; a third compared BAC and BACE groups. A modified Alexander trend analysis (**1**) was the principal technique for assessing the effects of the independent variables. The differences between the analyses used and that outlined by Alexander are that quadratic components of trend are considered in addition to slopes and means, and that Alexander does not elaborate his method for the required two-variable case where interactions must be assessed. Trial values were coded by orthogonal polynomial coefficients (**5**) in order to extend the analyses as far as quadratic components.[3]

Table 1 presents the results of the three analyses. In all three analyses the mean square for individual deviations from estimation was used as the error term for testing differences among individuals, over-all trend components, and group deviations from estimation, in short, all terms from *j* through *p* in Table 1. The finding that all of these terms were highly significant in each of the analyses is very reasonable and deserves only passing comment. Sig-

[3] This procedure was suggested by Dr. David A. Grant. The computational formulas were provided by Dr. Charles O. Nystrom.

nificance of these terms merely means that there were consistent individual differences, there was an over-all learning effect, and there were some higher (than quadratic) order differences between group trends.

The error term for testing any group differences is always the corresponding individual term, e.g., between group slopes mean squares are tested with between individual slopes as error. Considering first the B-BA analysis, Table 1 shows that variations in percentage of misinformative feedback introduce highly significant differences in group means, slopes, and quadratics. These differences are apparent in an examination of the B and BA curves in the subplots of Fig. 1 where negatively accelerated rapid acquisition is observed for the zero per cent groups, somewhat slower and more linear learning is observed for intermediate percentages, and no learning at all is found at some of the high percentages. One of the most important results is that awareness of misinformative feedback also had significant effects on group means and group slopes (term *b* and *e* in the B-BA analysis of Table 1). In Fig. 1 the superiority of the BA over the B group is especially marked with 10% misinformation but is also evidenced at the 20% and 30% levels. At 40% and 50% neither the B nor the BA curves seem to show any improvement with practice. In other words, the effect of awareness depends on percentage of misinformative feedback and a significant interaction between means is observed in Table 1 (term *c* in B-BA analysis).

The use of an unexplained correction signal in the feedback channel on misinformative trials is evaluated in the BA-BAC analysis of Table 1. The correction signal significantly improved performance, and percentage of misinformative feedback again introduced significant effects. Though the differences are sometimes small and hard to discriminate, it is true, nevertheless, that in every subplot of Fig. 1 the BAC curve has a greater mean, a steeper slope, and a greater quadratic component than the BA curve. The BAC curves even show some learning at the 40% and 50% levels.

The final analysis of Table 1 and a comparison of the BACE and BAC curves of Fig. 1 make it clear that an explanation of the correction signal still further improved performance, especially that of groups with high percentages of misinformative feedback. An analysis done but not presented compared the six BACE groups alone and there were no significant differences among these curves.

To this point no detailed consideration has been given to terminal levels of response accuracy and to whether these levels show that a particular response follows a stimulus about as often as that stimulus-response pair is rewarded. This has been called probability matching behavior. Before one can say that either an individual or a group curve demonstrates probability matching behavior it is important to know that the curve has reached a stable state and has a slope not significantly different from zero. If learning is still taking place the answer to the question remains indeterminate unless, of course, the probability matching level has already been surpassed. In Fig. 1 the percentage of correct responses that defines a probability matching solution is one minus the percentage of misinformative feedback given in the upper left of each subplot. Considering only the data of the B and BA groups with from 10% to 40% misinformative feedback in detail, only the 30% BA group seems to approximate matching behavior. The answer is indeterminate for the 10%

and 20% B' groups since both still have positive slopes and are below the matching levels at the termination of practice. The 10% and 20% BA groups, on the other hand, surpass a matching solution, having terminal response accuracies of 98.3% and 84.6%, respectively. During an even smaller terminal segment, the final 16 trials, the probability rises to 87% for the 10% BA group. With no learning in the other groups, they show a non-matching stable state.

One cannot tell from group curves whether individual curves show probability matching. Decisions about individual curves should take into account consistent individual differences, and the error for testing whether an individual curve shows matching behavior should be based on *S*'s variability, not on a theoretical standard error of a proportion for the matching probability. When the criteria described are applied in an inspectional analysis, then never more than three of the 12 *S*s in any of the groups under consideration showed probability matching responses. Though as with group curves there were cases of indeterminacy, for the most part the negative cases clearly did not show matching behavior.

It has been suggested that awareness of misinformative feedback might make *S*s less inclined to reject hypotheses which return less than 100% yield. This has been found true for the optimal solution hypothesis, but a question still unanswered is whether it is also true for nonoptimum hypotheses. Written comments at the end of the experiment indicated that a variety of incorrect but systematic modes of response had been used in attempts to solve the problem. It was decided to take a single but rather frequent one of these and examine it thoroughly. This was the tendency to persist in incorrect responses to a single stimulus light. For example,

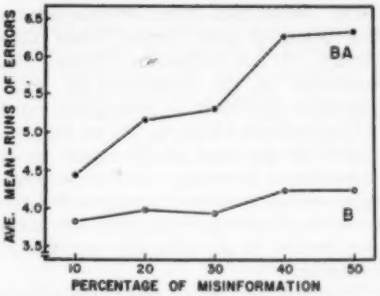


FIG. 2. Group averages where entry for each *S* is mean length of runs of three or more errors to single stimulus lights. Data for B and BA groups with 10% to 50% misinformative feedback.

if the left key was correct for a particular light, then runs of three or more right key responses to that light were counted. The question was: given that a run of *at least* three wrong responses has occurred, is it more likely that a run will persist when *S* is aware of misinformative feedback? For each *S* the counts provided a frequency distribution of runs from which a mean was computed to serve as the basic entry in an analysis of variance on all B and BA groups with 10% or greater misinformative feedback. In Fig. 2 average mean-runs of wrong responses is plotted as a function of percentage of misinformative feedback. At every percentage of misinformation mean-runs of errors are greater in the BA group. The *F* comparing awareness (BA) and nonawareness (B) was 63.03 and highly significant with 1 and 110 *df*. There was also a significant linear increase in the average mean length of runs as percentage of misinformative feedback increased. The *F* was 5.56 and was based on 1 and 110 *df*.

DISCUSSION

The adequacy with which *S*s can filter out misinformative feedback and acquire optimum response solutions has been shown to be a decreasing function of the percentage of misinformative feed-

back. With high levels of misinformation learning may break down completely. It has further been shown that awareness of the possibility of misinformative feedback, a correction signal in the feedback channel, and an explanation of the function of this signal are all important to learning. As these changes in procedure are cumulatively added to a set of basic instructions, each new addition results in significantly better performance.

The task with which these results have been found is different in several respects from that used in other studies of misinformative feedback, and some of these differences may be related to the results. For one thing, in earlier work *S*s have been required to make only a single frequency discrimination. The only stimulus for this discrimination was a series of rewards and nonrewards of the two choices. Optimum behavior always has been uniform selection of one of the two responses. This was changed to some extent in this study by requiring eight simultaneous frequency discriminations of *S*. It was necessary to pay attention to the series of rewards and nonrewards as they depended on which stimulus light had initiated the trial. So even when *S* reached optimum performance he was giving half of his responses to the left key and half to the right key. This complication plus appropriate instructions made the task more like a problem to be solved than like a gambling or guessing game. There was no indication in the written comments of any *S*s that indicated they felt they were guessing or predicting events. The task also probably minimized the operation of special value systems which some investigators (7) have felt may play a role in failures to find optimum behavior. In guessing or gambling studies, for example, it seems likely that *S*s place special value on predicting a rare event.

It was within the context of a problem-solving task, then, that awareness of misinformative feedback was shown to be an important variable. The *S*s given only basic learning instructions had no immediate reason to doubt that a completely deterministic solution was possible. Even among those *S*s who did reach optimal solution when there was 10% or more misinformative feedback, almost without exception those in the B groups indicated, in answer to a written question, that they had only partially solved the problem. Similar *S*s in the BA groups answered the same question by saying that they had completely solved the problem. Instructing BA groups about the possibility of the apparatus "lying" apparently made *S*s believe some errors were unavoidable, and consequently they were less inclined to reject a hypothesis which did not pay off on every trial. It is especially interesting that this resulted not only in more optimum behavior but also in greater persistence in one incorrect mode of response. Mean-runs of errors to single stimulus lights were on the average longer in the BA than in the B groups. A real problem for *S*s who believe correctly that some errors are unavoidable hangs on the difficulty of overthrowing solutions which give temporary chance success or permanent partial but nonoptimum success. As far as *S* is concerned any given avoidable error may belong to the set of unavoidable errors. An interesting question is whether this might ever result in fixation on nonoptimum hypotheses to the extent that optimum solution never occurs.

It is regrettable that practice was not carried far enough to determine for all groups whether probability matching behavior would eventually occur. The task proved especially difficult for the B groups. Several group curves, however, especially those of some BA groups, had either surpassed a matching solution or appeared to be describing a nonmatching stable state. Of prime importance in this problem was the finding that even when a group curve did approximate a matching solution, the behavior of most *S*s in that very group could not be similarly described. The criteria for deciding whether individuals show matching behavior have already been elaborated. The necessity of applying criteria this strict rests on the fact that fundamentally different explanations may be needed to account for the behavior if group curves show matching but individual curves do not. Though no

theory is offered to account for the observation of matching behavior in many investigations, the results of this study are important for such theories (**4, 7, 9**) since the results do indicate there are conditions under which matching behavior is not found.

The addition of an unexplained correction signal into the feedback channel (BAC groups) improved performance and a further improvement was obtained when both the signal and an explanation of its function were given (BACE groups). In one sense the term *misinformative feedback* was a misnomer when applied to the BACE groups. An explanation of the function of the correction signal redefined a correct response as one producing a green light, or a red light followed by the correction signal, and a wrong response as one producing a red light, or a green light followed by the correction signal. The redefinition made the problem relatively easy for most *S*s at every percentage of "misinformative feedback." In the BAC groups there were evidences of a self-instructed redefinition of success. Most *S*s in the BAC groups who solved the problem had surmised its function by the end of the practice period. This was shown in the written comments. Again *S*'s conceptualization of the task was found to be important to performance.

Summary

On each of 240 trials one of eight white stimulus lights arranged in a circle came on and *S* selected one of two response keys. Selection of the correct key lit a green feedback light and the wrong key lit a red feedback light *except* on misinformative trials. On these trials *S* was misinformed about the adequacy of his response and the opposite feedback light came on. The left key was correct (most often rewarded) for four stimulus lights and the right key was correct for the other four. Groups were given 0, 10, 20, 30, 40, or 50% misinformative feedback. A second categorization depended on instructions and the presence or absence of a correction signal. The principal findings were as follows:

1. When *S*s were aware of misinformative feedback they learned significantly faster and more went to optimal solution. The effect was most pronounced at 10% misinformation but was also apparent at the 20% and 30% levels.

2. Use of an unexplained correction signal near the end of misinformative trials significantly improved performance.

3. A further significant improvement was obtained when the correction signal was explained.

4. In general, the difficulty of learning optimal behavior increased as the percentage of misinformative feedback increased.

5. Considering runs of at least three consecutive errors to a single stimulus light, the average length of such runs was significantly greater when *S*s were aware of possible misinformative feedback. In other words, awareness also increased persistence in incorrect modes of response.

6. Most individuals and groups were not showing probability matching behavior at the termination of practice. In some cases this may have been because practice was not carried far enough, but frequently it was because individuals or groups had surpassed a probability matching solution or reached a stable state below the matching level.

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. Brunswik, E. Probability learning as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–177.
3. Edwards, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.
4. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
5. Fisher, W. K., & Yates, F. *Statistical tables for biological, agricultural and medical research.* (4th Ed.) Edinburgh: Oliver and Boyd, 1953.
6. Grant, D. A. Information theory and discrimination of sequences in stimulus events. In R. A. Patton (Ed.), *Current trends in information theory.* Pittsburgh: Univ. of Pittsburgh Press, 1953.
7. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
8. Humphreys, L. G. Acquisition and extinction of a verbal conditioned response in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
9. Jarrett, J. M. Strategies in risk-taking situations. Unpublished doctor's dissertation, Harvard Univ., 1951.
10. Jarvik, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

(Received June 29, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

TRANSFER OF SECONDARY REINFORCEMENT ACROSS THE HUNGER AND THIRST DRIVES [1]

MICHAEL R. D'AMATO

New York University

The purpose of the present studies is to demonstrate that a neutral stimulus correlated with one set of motive-incentive conditions (e.g., hunger-food) will act as a secondary reinforcer in a T maze when motivation is shifted to an inappropriate drive (e.g., thirst). Such transfer of secondary reinforcement across drives bears directly on the question of the role of drive reduction in learning. For if it can be shown that a stimulus can reinforce behavior motivated by drives for which the reinforcing stimulus can have no relevance, either as a primary or a secondary reward, then either drive reduction must be abandoned as a necessary condition for all learning, or else different mechanisms must be advanced to accommodate primary and secondary reinforcement (**5, 6**).

It has already been demonstrated that satiated rats will (*a*) prefer the side of a T maze on which food and water appear (**8, 9**), and (*b*) run faster on a straight-away to an end box that contains food than to an empty end box (**13**). In addition, thirsty rats have been shown to run faster to an end box containing food than to an empty one (**12**). One may interpret these results to mean that the sight and smell of food (and sight of water)—secondary reinforcers—can strengthen behavior when the appropriate drive is supplanted by an inappropriate one (**12**), or when no experimentally manipulated drive is present (**8, 11, 13**). In the first case, while it seems clear that the reinforcing stimuli have no relevance for the dominant drive (thirst) as far as primary reward is concerned, it is possible that the stimuli have relevance for thirst as secondary rewards, owing to the prior close association between food and water, and eating and drinking. However, where food and water act as reinforcing stimuli in the face of *no* manipulated drive the situation is somewhat different. Let us assume the presence of an "exploratory" drive. Under such an assumption it seems equally clear that the mere *sight* of food and water can have no relevance as "primary" rewards for the exploratory drive; and, considering the usual history of the laboratory rat, it also appears unlikely that such stimuli can claim relevance as secondary rewards. Thus, the question raised in the opening paragraph would seem to have already been answered.

However, since food pellets "scattered on the goal-box floor" and mash and water in cups are manipulable objects, the question arises as to whether the rats of the above studies were running to Purina pellets (or mash) as *food* objects or as objects of manipulation. In other words, the sheer presence of manipulable objects must be controlled before we may safely conclude that *S*s were running to food (and water) rather than to small, oblong, dark objects, etc. In the latter case, a relevance can be claimed between the reinforcing stimuli (manipulable objects) and a component of the concurrent motivation (exploration). Wike and Casey (**13**) have recently acknowledged this point.

Seward and Levy (**11**, Exp. II) sought to evaluate this variable by having, for half the *S*s, food and water in *both* the

[1] These studies, which were briefly reported in a paper read at the 1954 convention of the APA, comprise part of a dissertation submitted to the faculty of the Graduate School of Arts and Science of New York University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to express his gratitude to Professor H. H. Kendler for his helpful guidance.

former positive and negative end boxes during testing for secondary reinforcement; empty food and water cups were in *both* end boxes for the remaining *S*s. Under these conditions it is rather remarkable that they obtained positive results. In any event, since they used the somewhat questionable method of giving *S*s differential training to the two end boxes, it is not clear whether, on the test trials, *S*s were going *to* the end box in which they had been previously rewarded, or whether they were going *away* from the end box in which they had previously encountered nonreinforcement (4). A procedure that correlates one cue with reinforcement and another with nonreinforcement and then pits the two cues against one another unavoidably confounds the effect of the two stimuli. In the present experiments a partial solution was achieved by having *S*s choose, on the test trials, between the former goal and starting boxes. Such a method (*a*) enables *E* to equate both the time spent in the two boxes and, although not as perfectly, *S*s' prevailing motivation while in the boxes; (*b*) at the same time the starting box, being almost equally associated with reward and nonreward (a 60% reinforcement schedule was used), probably develops little, if any, negative (avoidance) properties for *S*s. The question of the "exploratory" value of the secondary reinforcing stimuli was largely avoided by using an empty reward cup, securely anchored to the goal-box floor, for this purpose.

Experiment I

Method

In this experiment a neutral cue is correlated with water reward and subsequently tested as a secondary reinforcer with *S*s' drive shifted to hunger.

Subjects.—The *S*s were 20 experimentally naive, male albino rats, 60–90 days of age at the start of the study.

Apparatus.—The apparatus consisted of two identical end boxes and a small unpainted T maze of the following dimensions: height, 14 cm.; width of stem and arms, 11 cm.; length of stem, 30 cm.; length of arms, 15 cm. Both arms terminated into guillotine doors. The end boxes were painted dark gray, and a white card, 10 × 15 cm., could be attached to the rear wall by means of a pair of grooves. The boxes were 15 cm. high, 15 cm. wide, and 30 cm. long. Each end box was equipped with a movable floor section, a slight depression of which turned on a red light.

Pretraining.—Six days prior to the beginning of training *S*s were segregated and the following motivational schedule commenced: H, T, H, T, T. On "H" days *S*s were 23 hr. hungry; on "T" days, 23 hr. thirsty. Throughout this and the following experiment, food was in the nature of Purina chow pellets, and water was obtained in the home cages from the tube of a water bottle.

Training.—All *S*s were given eight trials per day for five days on the straightaway formed by the arms of the T maze, the stem being removed. Goal and starting boxes differed in that the goal box contained a white water cup while the starting box had the white card attached to the rear wall.[2] The end boxes alternated daily as starting and goal box.

A 60% reward (R) schedule was employed, with the following distribution: Day 1, RRRNRNRN; Day 2, RRNNRRNR; Day 3, RNRRNRNR; Day 4, RRNNRNRR; Day 5, RRNNRNNR. The running procedure was as follows. Groups of six or seven *S*s were transported into the experimental room in the individual compartments of a carrying cage; there was food in each compartment. The *S* was placed in the starting box, and 15–20 sec. later the door was raised. When *S* entered the goal box, the door was lowered behind him, and he was permitted to drink for 15–20 sec.; on nonrewarded trials *S* remained in the goal box with the empty water cup for 15–20 sec. After the daily trials, *S*s were returned to the home cages where, 20–30 min. later, they received their water bottles for 1 hr.

Prior to the trials of Training Day 4, gray curtains were attached behind the guillotine doors of both starting and goal box, so that in order to leave the starting box or to enter the goal box *S* had to push through the curtain.

On the average, *S*s started the daily training session 21.8 hr. thirsty; range: 21.5–23.0 hr. The average intertrial interval for Trials 2–40 was 8.1 min.; range: 5–17 min.

Testing.—Upon termination of the feeding period of Training Day 5, food only was removed from the home cages. The next day, 20–30 min.

[2] A prior study (4) showed that the location of the card in either the starting or goal box was an unimportant factor. The present procedure was employed for reasons of experimental simplicity and as an effort to balance the stimulus dissimilarity between the end boxes.

prior to testing, a flat dish of water was introduced into the cages. During testing, in addition to this same water dish, a water tube was present in each carrying cage compartment.

Testing consisted of 15 free choices on the T maze previously described. The stem of the T was oriented along the same axis as had been occupied by the straight alley during training. The starting and goal box of training served as end boxes. Thus, the assumption was that if, under the altered drive conditions, the former goal box possessed any secondary reinforcing strength, *Ss* should respond to it significantly in excess of chance expectancy. For the sake of convenience, a turn to the side on which the former goal box is located will be called a "correct" response; a choice of the arm leading to the former starting box is an "incorrect" response. For a particular *S*, the position of the former goal box was determined as follows: Half the *Ss* were randomly preassigned to be "correct" on Test Trial 1. On the first test trial these *Ss* found the former goal box whichever way they turned. The reverse was true for *Ss* preassigned to be "incorrect" on Test Trial 1; they found the former starting box at the end of whichever arm they chose. This procedure assures a 50% baseline on the initial test trial, an important control.

During testing, *S* was placed directly into the stem of the maze; after making his choice, the guillotine door was lowered behind him, and *S* remained in the end box for 15–20 sec. A choice was defined as an entry into an end box sufficient to activate the floor panel and thus cause the red light to signal. The curtain and floor panel arrangement was such that *S* tripped the floor switch upon thrusting his head through the curtain.

The average hunger drive at the beginning of testing was 21.8 hr.; range: 21.5–22.0 hr. The average intertrial interval was 7.7 min.; range: 5–16 min.

Results and Discussion

The mean number of responses to the side of the T maze on which the former goal box was located was 8.80 ($SD = 1.47$) for the 20 *Ss*. This figure exceeds chance expectancy (7.50) at the .002 level ($t = 3.855$). It appears, therefore, that a secondary reinforcer set up under thirst motivation and water incentive will generalize to the hunger drive even when the relevance of the reinforcing cue for the exploratory drive is minimized and prior association of this stimulus with the hunger drive is precluded. In the first row of Table 1, the trial by trial percentage of "correct" responses is presented.

These results seemingly contradict data previously reported by the present writer (3). That study, very similar to the present one, failed to produce transfer of secondary reinforcement. The discrepancy can probably be attributed to the following factors: (*a*) In the earlier experiment the position of the goal box during testing was randomly assigned, and quite inexplicably 9 of the 10 *Ss* tested responded "incorrectly" on the first test trial. Randomization, therefore, failed to control adequately for the possible effects of position preferences. (*b*) Although 15 test trials were planned, a few *Ss* lingered so long in their choices that it was necessary to truncate the test period to 12 trials. If the results of the 10 *Ss* of the present study that were "incorrect" on Test Trial 1 are analyzed for the first 12 trials, it is found that their performance does not significantly exceed chance expectation. Hence, the earlier failure to find transfer of secondary reinforcement can with reasonable

TABLE 1

PERCENTAGE OF RESPONSES TO FORMER GOAL BOX

Experiment	Trials														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
I	50	60	50	65	55	50	65	65	75	45	60	40	60	60	80
II	50	70	30	40	85	45	60	55	60	75	60	55	70	55	60

confidence be assigned to the joint operation of these two factors.

Experiment II

In this experiment, the symmetrical obverse to Exp. I, *Ss* are trained hungry and tested while thirsty. The only data relevant to this problem are those reported by Wike and Casey (**12**). They used a straight alley and employed food pellets, scattered on the goal-box floor, as the secondary reinforcing stimuli.

Method

Subjects.—The *Ss* were 20 experimentally naive, male albino rats, 60–90 days of age at the start of the study.

Apparatus.—The same apparatus used in the previous experiment was employed.

Pretraining.—Motivational training lasted six days as follows: T, H, T, H, H, H. On T days, *Ss* were 23 hr. thirsty; on H days, 23 hr. hungry.

Training.—The training procedure, exclusive of reward, was identical to that of Exp. I. Reward consisted of three very dry pellets, each weighing about .063 gm. Water was always available to *Ss* in the carrying cage compartments.

The average strength of the hunger drive at the commencement of a day's trials was 21.5 hr.; range: 20–23 hr. The average intertrial interval was 8.7 min.; range: 5–16 min.

Testing.—At the end of the 1-hr. feeding period of Training Day 5, water but not food was removed. The next day, 20–30 min. prior to testing, a small dish of reward pellets was introduced into the home cages. All *Ss* were observed to eat some of these dry pellets, a few eating large numbers of them.

During the 15 test trials reward and Purina pellets were available in the carrying cage compartments. The test trials were otherwise conducted as those of the previous experiment.

The average duration of water deprivation at the start of the test trials was 21.8 hr.; range: 21.5–22.0 hr. The average intertrial interval was 8.9 min.; range: 5–24 min.

Results

The present results closely parallel those of Exp. I. The mean number of responses to the "correct" side was 8.70 ($SD = 1.10$). This figure exceeds chance expectancy at better than the .001 level of confidence ($t = 4.754$). The second row of Table 1 shows the percentage of correct responses for each trial.

Discussion

In attempting to account for the results of the present experiments, faulty satiation of the unwanted drive must be ruled out as an important factor because of the presence of the appropriate commodity, food or water, in the carrying cage compartments throughout testing.[3] For the same reason, appeal to conditioned primary drive as providing a possible source of drive reduction must be denied (**1, 2, 5, 9**). It therefore becomes difficult to account for the present results in terms of drive reduction. If we take into consideration the presently available data (**4, 6, 7, 8, 11, 12, 13**), a fair generalization of the conditions under which a secondary reinforcer operates might be as follows: A secondary reinforcer established under one set of motive-incentive conditions will reinforce any response which it follows, in general the degree of reinforcement being maximal when concurrent motivation is appropriate (i.e., when the drive state during testing is the same as prevailed during the establishment of the secondary reinforcer).

Apparently, the only solution presently available to those holding to the drive-reduction interpretation of secondary reinforcement seems to be in a development of Mowrer's somewhat implausible notion of "drive-fear" (**10**). The writer has elsewhere reviewed the entire problem and come to the conclusion that the drive-reduction interpretation of secondary reinforcement cannot at present account for a rather large body of relevant data (**4**).

[3] An unpublished study (**4**) has established that the presence of the appropriate commodity in the carrying cage compartments during training is not a necessary condition for transfer of secondary reinforcement.

Summary

In Exp. I, 20 rats were given eight trials per day for five days on a straightaway, *Ss* being thirsty and receiving a water reward. Subsequently *Ss*, while hungry, were given 15 free choices in a T maze to determine whether the former goal box possessed secondary reinforcing strength under the altered motivating conditions. The *Ss* showed a highly significant preference for the side of the T maze on which the former goal box was located. In Exp. II *Ss* were trained hungry and tested for secondary reinforcement while thirsty. Secondary reinforcement was again in evidence.

References

1. Anderson, E. E. The externalization of drive. III. Maze learning by non-rewarded and by satiated rats. *J. genet. Psychol.*, 1941, **59**, 397–426.
2. Calvin, J. S., Bicknell, E. A., & Sperling, D. S. Establishment of a conditioned drive based on the hunger drive. *J. comp. physiol. Psychol.*, 1953, **46**, 173–175.
3. D'Amato, M. R. Drive shift and secondary reinforcement. *Amer. Psychologist*, 1953, **8**, 338–339. (Abstract)
4. D'Amato, M. R. A study of the role of the relationship between concurrent motivation and antecedent drive-incentive conditions in secondary reinforcement. Unpublished doctor's dissertation, New York Univ., 1954.
5. Danziger, K. The operation of an acquired drive in satiated rats. *Quart. J. exp. Psychol.*, 1951, **3**, 119–132.
6. Estes, W. K. Generalization of secondary reinforcement from the primary drive. *J. comp. physiol. Psychol.*, 1949, **42**, 286–295.
7. Estes, W. K. A study of motivating conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1949, **39**, 306–310.
8. MacCorquodale, K., & Meehl, P. "Cognitive" learning in the absence of competition of incentives. *J. comp. physiol. Psychol.*, 1949, **42**, 383–390.
9. Meehl, P., & MacCorquodale, K. Drive conditioning as a factor in latent learning. *J. exp. Psychol.*, 1953, **45**, 20–24.
10. Mowrer, O. H. Comment on Estes' study: "Generalization of secondary reinforcement from the primary drive." *J. comp. physiol. Psychol.*, 1950, **43**, 148–151.
11. Seward, J. P., & Levy, N. Choice-point behavior as a function of secondary reinforcement with relevant drives satiated. *J. comp. physiol. Psychol.*, 1953, **46**, 334–338.
12. Wike, E. L., & Casey, A. The secondary reinforcing value of food for thirsty animals. *J. comp. physiol. Psychol.*, 1954, **47**, 240–243.
13. Wike, E. L., & Casey, A. The secondary reward value of food for satiated animals. *J. comp. physiol. Psychol.*, 1954, **47**, 441–443.

(Received July 16, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

EFFECT OF MONOCULAR AND BINOCULAR VISION, BRIGHTNESS, AND APPARENT SIZE ON THE SENSITIVITY TO APPARENT MOVEMENT IN DEPTH

WILLIAM M. SMITH

Princeton University

Previous experiments (**12, 13**) have reported the effect of certain factors upon the perceptual sensitivity to apparent movement in depth. This type of apparent movement is observed when a stimulus, viewed against a homogeneous background, is changed continuously in angular size. As angular size is increased, the stimulus appears to move toward *S*; as angular size is decreased, it appears to move away. In the experiments cited above, sensitivity (as measured by reaction time) to this kind of illusory movement was studied in relation to particular apparent or meaningful properties of stimuli. Results showed that sensitivity did not vary significantly with objects which normally are perceived as moving, e.g., baseball, and those which are not, e.g., a geometrical form. Furthermore, it was found that sensitivity was no greater for stimuli that appeared three-dimensional than for those that appeared two-dimensional. However, it was found in one of these studies (**13**) that sensitivity did vary significantly with the mode of observation, binocular vision favoring greater sensitivity. This result was somewhat unexpected because this sort of illusory movement always has been considered as essentially a monocular phenomenon. One purpose of the present experiment was to repeat observations on this monocular-binocular problem in order to confirm or deny the findings of the previous experiment.

An additional purpose of the present experiment was to investigate the sensitivity to this kind of apparent movement as a function of brightness and apparent size. In view of data concerning other types of visual movement, it was expected that sensitivity to this type of movement also would vary with brightness. It was considered appropriate to investigate the relation between sensitivity and apparent size because the results of a previous experiment (**12**) suggested that sensivitity might be affected by phenomenal or apparent distance. By virtue of the relation between apparent distance and apparent size, it was hoped that more definitive information could be obtained on this point by determining possible variations in sensitivity as a function of apparent size.

METHOD

Apparatus and stimulus materials.—The apparatus was the same as that employed previously (**12, 13**) and a detailed description and diagram of it can be obtained there. The main component of the apparatus was a Clason projector whose movable lens element was driven by a planetary gear device through a rack and gear system. By this means the projected image could be varied continuously in angular size. The stimulus, a white equilateral triangle, was projected on a screen and then reflected by means of a front surface mirror and an evaporated metal film into *S*'s eyes. A cardboard mask placed between the front surface mirror and the evaporated metal film was used to occlude when desired the visual field of one eye. The *S* sat in a cubicle from which he made the observations and on top of which was located the Clason. A head rest was provided. In the

cubicle and to the right of *S* was located a response key which was used to indicate the detection of movement. This key was wired into the circuit so that when it was pressed the projector light source and power were turned off and the Standard Electric Timer was stopped. When *E*'s stimulus key was pressed the projector was turned on, and the transmission and timer were started.

On the wall of the cubicle and to *S*'s right was placed that part of the apparatus used to measure the apparent size of the projected stimulus. A large, white equilateral triangle was suspended on a thin steel wire and could be moved by *E* up and down behind the top edge of a flat black panel. As the triangle was moved up more of its surface was exposed; as it was moved down its size decreased. The suspension wire led to a scale mounted on the outside of the cubicle. Any given position of the triangle could be converted directly to a linear size measure by reading the scale to the nearest .125 in. This comparison triangle was illuminated by a diffusion-type light source and its brightness was 1 ft.-L. as measured by a Macbeth illuminometer.

The brightness of the stimulus was controlled by a Variac. The change in voltage caused a slight change in the hue of the stimulus, but this was judged by *E* to be so slight that no attempt was made to correct for this optically.

The distance of the stimulus from *S*'s eyes was 12 ft. 8 in. Its size (side) was 51 mm. These values correspond to an angular size of 45.5 min. The stimulus was changed in size (linear) at a rate of 2 mm./sec. The angular velocity was thus approximately 1.5 min./sec. The stimulus was always viewed against the black background afforded by the dark experimental room.

Experimental design.—The design of the experiment required the measurement of reaction time to apparent movement in depth for both monocular and binocular vision under two brightness conditions, .2 and 2 ft.-L. In using reaction time as a measure of sensitivity, it was realized, of course, that the time of the motor response itself was confounded with the time of perception. Both Klein (**8**) and Gordon (**3**) have discussed this problem. In this experiment it was assumed that the motor component of the reaction time was reasonably constant for the different experimental conditions employed. Measures of the apparent size of the stimulus object for all experimental conditions also were obtained.

In order to utilize all possible sequences of the four experimental conditions of observation, 24 high school students with 20/20 vision, corrected or uncorrected, served as *S*s. All *S*s were naive with respect to the purpose of the experiment.

Procedure.—Each *S* was dark-adapted for 5 min. in the cubicle and was then shown the projected stimulus under the experimental condition assigned first in his particular sequence of observations. He was instructed to observe carefully the triangle which appeared directly in front of him, and then the comparison triangle to his right (3-ft. distance) whose size was to be changed until it appeared equal to the triangle directly ahead. Equality of size was defined for *S* by stating that when he judged the two triangles to be equal, they should be of the same linear dimensions if placed side by side. A method of adjustment was used, although *E* varied the size of the comparison triangle according to directions from *S*. Ascending and descending series were employed in alternate order.

After the measure of apparent size was obtained, the light in the cubicle was turned off, and *S* was given the following instructions.:

"All that you see now is the object directly ahead of you. What I want you to do now is watch this object carefully and, at the exact instant you see it move toward you, press the key on your right. Always keep your arm and hand resting in position on the key mount. Remember, press the key at the exact instant you see the object begin to move toward you; not when you *think* it begins to move, but when you *see* it begin to move. After you have pressed the key, relax and wait for instructions concerning the next trial. Do you have any questions?"

Three more practice trials were given, one for each of the three remaining conditions of observation. At the end of the practice trials questions of procedure were clarified before continuing.

Four trials on each of the four experimental conditions were obtained from each *S*. All trials under one condition were made before proceeding to the next condition. The measurement of apparent size was made prior to each trial. "Catch" trials, i.e., no change in size, but motor noise, etc., were randomly inserted among trials in order to help control anticipation. Between trials an occluding mask was dropped between *S*'s face and the viewing box while *E* prepared for the next observation. At the end of his series of observations *S* was questioned briefly about certain aspects of the experiment.

Results

Reaction time data.—The measures of reaction time to apparent movement in depth were made to the nearest .01

TABLE 1

REACTION TIMES IN SECONDS FOR THE FOUR EXPERIMENTAL CONDITIONS
(N = 24 each condition)

Vision	Brightness			
	.2 ft.-L.		2 ft.-L.	
	Mean	*SD*	Mean	*SD*
Monocular	5.46	2.41	4.22	2.05
Binocular	4.66	2.15	3.41	1.49

sec. The general magnitude of these measures was much greater than usual reaction time measures, as Table 1 shows. Each mean given in the table is based upon 96 observations, 4 from each of 24 *S*s. The effect of viewing condition and brightness on reaction time is indicated by the means given in the table. As brightness increases, the reaction time to apparent movement in depth decreases, as it does when observation is changed from the monocular to the binocular condition.

In order to analyze the time data all measures were transformed to common logarithms. The four trials for each *S* under each condition first were summed to give a mean log score. The data in this form also indicated proportionality between means and variances apparent in the raw data of Table 1. For this reason a test of significance of the equality of variances of correlated observation was made on the transformed data using Wilks' L_{ve} criterion (**15**).[1] The two most different distributions (monocular low brightness vs. binocular high brightness) only were tested. The value of L_{ve} derived from these data was .8571 which is above the 5% probability point for L_{ve} (.8365) (**15**, p. 263, Table 1) for $k = 2$ and $N = 24$; hence, it can be safely concluded that the variances are not significantly different between the two conditions, and by inference, that the variances among all conditions are not significantly different. On the basis of this interpretation, an analysis of variance was then carried out on the transformed data. The results of this analysis are given in Table 2.

First-order interactions were tested against the triple interaction term and, as the table shows, two of the three are significant at the .01 level. Both involve the factor of individual differences. When the main effects of brightness and viewing conditions are tested against their appropriate interaction terms (1×3 for brightness and 2×3 for viewing condition), both F values are significant, at the .001 level for brightness and the .01 level for viewing condition. The presence of the two significant first-order interactions, both of which involve variance due to *S*s, makes it impossible to test directly the main effect of *S*s. The mean square for this main effect is large, however, and differences among individuals is ob-

TABLE 2

ANALYSIS OF VARIANCE OF MEAN LOG REACTION TIME TO APPARENT MOVEMENT IN DEPTH

Source	*df*	Mean Square	*F*
1. Brightness	1	.3504	26.95**
2. M-B vision	1	.1569	13.76*
3. *S*s	23	.1146	—
4. 1×2	1	.0001	0.03
5. 1×3	23	.0130	3.71*
6. 2×3	23	.0114	3.26*
7. $1 \times 2 \times 3$	23	.0035	—
Total	95		

* Significant at the .01 level.
** Significant at the .001 level.

[1] A similar test of equality of correlated variances is described by Kenny (**7**). This test was also applied to the data mentioned above. An $R = .24$ was obtained which gave a t value of 1.16 with 22 *df* and is not significant.

TABLE 3

APPARENT SIZE IN INCHES FOR THE FOUR CONDITIONS (N = 24 each condition)

Vision	Brightness			
	.2 ft.-L.		2 ft.-L.	
	Mean	*SD*	Mean	*SD*
Monocular	2.57	1.33	2.71	1.32
Binocular	2.65	1.27	2.77	1.26

viously a major source of variance. Notice also in the table the size of the mean square for brightness which is larger than all other sources of variance combined. The results of this analysis clearly indicate that both variables of brightness and viewing condition affect the sensitivity of response to this type of apparent movement. The effect of these factors seems to vary from S to S, however, as the analysis of the first-order interaction shows.

Apparent size data.—Prior to each reaction-time trial, a measure of the apparent size of the stimulus was obtained. A total of 16 such measures was obtained from each S, 4 on each of the 4 experimental conditions. For each S and each condition, a mean apparent size was computed. The grand means and *SD*'s of the distributions of these individual means are given in Table 3.

In order to determine the relation between reaction time and apparent size, a rank-difference correlation was computed for each experimental condition separately. The coefficients for the monocular low-brightness, monocular high-brightness, binocular low-brightness, and binocular high-brightness conditions were .29, .08, .23, and .03, respectively. None of these values approaches significance, even at the .05 level (**14**). The speed with which S responded to the apparent movement in depth was not related to the perceived size of the stimulus.

Although not directly relevant to the main purpose of the experiment, another aspect of the data on apparent size deserves mention. As Table 3 shows, apparent size varied with both brightness and viewing condition. The means are larger for both binocular vision and the higher brightness level. An analysis by t tests of the differences among the means given in Table 3 reveals that only the most extreme difference is significant—that is, the difference between the monocular low-brightness and the binocular high-brightness condition. The t value here is 2.52 which is significant between the .01 and .02 level with 23 *df*. Neither a change in brightness nor a change in viewing condition alone significantly altered the apparent size of the stimulus, but both factors operating together did.

DISCUSSION

Present results concerning mode of vision confirm previous findings (**13**) which indicated greater sensitivity for binocular vision. The superiority of binocular vision for this type of visual discrimination is in line with other types of monocular-binocular differences in sensitivity (**4, 5, 6, 10**). The binocular superiority found here does raise a question, however, with regard to the specific phenomenon of apparent movement in depth. This question was brought up in the previous study (**13**) because the phenomenon usually has been presented as a monocular one, the implication being that the effect depended upon inadequate distance perception—a condition not met when binocular vision is employed. Binocular vision per se does not insure adequate perception of distance, however, and probably added little, if any, distance information in the present ex-

periment, since disparity as a factor was ruled out by using two-dimensional stimuli. If it is assumed that binocular vision in this experiment added little to correct distance perception of the stimulus and, furthermore, if it is realized that *S*s were judging under specific instructions to perceive movement, it is not surprising that greater sensitivity was found for binocular vision. Two eyes are apparently better than one when it comes to discriminating a change in size regardless of what the perceived effect of such a change might be.

The expectation of an increase in sensitivity due to an increase in brightness was confirmed. A difference in one log unit of brightness resulted in approximately a 25% decrease in reaction time for both monocular and binocular vision. The results for brightness are also in line with other visual sensitivity data which show a positive relation between brightness and sensitivity, at least up to a certain point (**1, 2, 9, 11**). The results of this experiment suggest that perhaps the change in angular size or a cue to depth or distance is affected more by the factor of brightness than by viewing condition. The group reaction times were reduced proportionately more in going from the lower to the higher brightness than changing from monocular to binocular vision. Significant interactions between each of these factors and *S*s indicate, of course, that not all *S*s were influenced in the same way by these two factors.

The relative importance of these two variables can be shown more clearly by considering the data in a different form. A kind of displacement threshold can be computed from the data by multiplying the angular velocity of the stimulus (1.5 min./sec.) by the reaction time. With the mean reaction time for the group of 24 *S*s under each condition, the displacement threshold turns out to be 8.2 min. for the monocular low-brightness condition, 7.0 min. for binocular low-brightness, 6.3 min. for monocular high-brightness, and 5.1 min. for the binocular high-brightness condition. From the data in this form one can say that changing from monocular to binocular vision reduced the displacement threshold by approximately 1 min. of angle, while raising the brightness of the stimulus reduced this threshold by approximately 2 min. of angle. These data agree with those of Basler (**1**), who reports a lower displacement threshold for high illumination than for low.

It is evident from the results that the apparent size of the stimulus had no effect on *S*'s sensitivity to movement. Related experiments (**12, 13**) have reported that other apparent or "nonstimulus" factors had no effect upon *S*'s sensitivity to this type of illusory movement. It would seem, at least so far as illusory depth movement is concerned, that sensitivity is affected by specific changes in stimulation, but not by apparent or functional properties of the perceived object. The suggestion in one previous experiment (**12**) that perhaps sensitivity here would be related to apparent distance is not supported, if the present apparent-size data are taken as presumptive evidence of changes in apparent distance. Apparent size and apparent distance vary together with constant retinal size. Since no relation was found between sensitivity and apparent size, it is reasonable to suppose that the same holds true for apparent distance.

Taking the present results for apparent size in conjunction with the results of two previous and related experiments (**12, 13**), all seem to be in disagreement with the current and popular hypotheses that certain "behavioral" or "nonstimulus" factors act to sensitize *S*'s perceptual response system.

Summary

The effects of brightness, monocular and binocular vision, and apparent size on sensitivity to apparent movement in depth were determined. Reaction time was used as the measure of sensitivity. Based upon an analysis of variance of the reaction-time data, and rho correlations determined between measures of apparent size and reaction time for each experimental condition taken separately, the following conclusions were reached.

1. Both brightness and mode of vision affect sensitivity to illusory movement in depth. Greater sensitivity was associated with the higher brightness value and binocular vision. The brightness factor seemed to affect sensitivity more than did mode of vision.

2. The apparent size of the stimulus was not related to *S*'s sensitivity to movement.

3. Interactions of *S*s with both factors of brightness and viewing condition were found in the data.

4. The findings of the experiment were related to other visual sensitivity data and to experiments dealing with the effects of so-called "behavioral" factors on perceptual sensitivity.

References

1. Basler, A. Über das Sehen von Bewegungen. I. Die Wahrnehmung kleinster Bewegungen. *Arch. ges. Physiol.*, 1906, **115**, 582–601.
2. Berry, R. N., Riggs, L. A., & Duncan, C. P. The relation of vernier and depth discriminations to field brightness. *J. exp. Psychol.*, 1950, **40**, 349–354.
3. Gordon, D. A. The relation between the thresholds of form perception, and displacement in parafoveal and peripheral vision at a scotopic level of illumination. *Amer. J. Psychol.*, 1947, **60**, 202–225.
4. Hirsch, M. J., & Weymouth, F. W. Distance discrimination: $\bar{V}$ Effect of motion and distance of targets on monocular and binocular distance discrimination. *J. aviat. Med.*, 1947, **18**, 594–600.
5. Horowitz, M. W. An analysis of the superiority of binocular over monocular visual acuity. *J. exp. Psychol.*, 1949, **39**, 581–596.
6. Ireland, F. H. A comparison of critical flicker frequencies under conditions of monocular and binocular stimulation. *J. exp. Psychol.*, 1950, **40**, 282–286.
7. Kenny, D. T. Testing of differences between variances based on correlated variates. *Canad. J. Psychol.*, 1953, **7**, 25–29.
8. Klein, G. S. The relation between motion and form acuity in parafoveal and peripheral vision and related phenomena. *Arch. Psychol.*, 1942, No. 275.
9. Mueller, C. G., & Lloyd, V. V. Stereoscopic acuity for various levels of illumination. *Proc. Nat. Acad. Sci., Washington*, 1948, **34**, 223–227.
10. Pirenne, M. H. Binocular and uniocular thresholds of vision. *Nature*, 1943, **152**, 698–699.
11. Shlaer, S. The relation between visual acuity and illumination. *J. gen. Physiol.*, 1937, **21**, 165–188.
12. Smith, W. M. Sensitivity to apparent movement in depth as a function of "property of movement." *J. exp. Psychol.*, 1951, **42**, 143–152.
13. Smith, W. M. Sensitivity to apparent movement in depth as a function of stimulus dimensionality. *J. exp. Psychol.*, 1952, **43**, 149–155.
14. Underwood, B. J., Duncan, C. P., Taylor, J. A., & Cotton, J. W. *Elementary statistics*. New York: Appleton-Century-Crofts, 1954.
15. Wilks, S. S. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math. Statist.*, 1946, **17**, 257–281.

(Received July 19, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

CHRONIC EFFECTS OF A HIGH-FREQUENCY STIMULUS ON THE STRUCTURE AND FUNCTION OF THE COCHLEA [1]

IRVING E. ALEXANDER AND FREDERICK J. GITHLER

Princeton University

The structural and functional changes that take place in the cochlea after intense high frequency stimulation are of some concern to the auditory theorist. Some of the more prominent bits of evidence supporting a "place" theory of pitch perception come from the post-mortem clinical material in which basal lesions are sometimes seen in old people who have suffered high tone hearing losses (**4**). Such evidence has been very difficult to reproduce experimentally.[2] This has been especially true when overstimulation has been the method of producing injury. In many cases lesions are produced in limited regions of the cochlea but the loss in sensitivity spreads throughout the audiogram (**8**). In other cases, despite losses in cochlear potentials, no gross changes in cochlear structures are observed (**7**).

Recently the present authors have suggested some changes in the methodology of stimulation deafness experiments in the hope of obtaining results that would be more reliable and help clarify the picture regarding the role of the cochlea in the analysis of pitch (**3**).

The present experiment, utilizing these methodological changes (control of magnitude of initial loss in sensitivity and observation of changes in a more chronic stage), was designed to measure the effects of a 10,000-cycle stimulus upon the electrophysiological responses of the ear and upon the cochlear structures themselves.

[1] This research was supported in part by the Office of Naval Research under Contract N6-ONR-270, T.O. 3, Project NR 140-322, and also by Higgins Funds allotted to Princeton University.

[2] Recently Schuknecht and Sutton (**5**) have published a paper in which abrupt high tone hearing losses were observed when basal lesions were introduced mechanically.

Procedure

Guinea pigs were used as *S*s. Using a sterile technique, a lateral approach to the auditory bulla was made and cochlear potentials were read from the round window membrane. A more detailed report of the procedure may be found elsewhere (**1, 8**). A 1-microvolt response audiogram was determined for the experimental ear. This is a plot of the sound pressures needed to produce a 1-microvolt response at the various frequencies. The ear was then stimulated with a pure tone of 10,000 cycles at a sound pressure of 1,000 dynes/cm.[2] The sound was maintained long enough to induce a decrement of approximately 40 db in the sensitivity to the exposure tone. When this decrement had been established the sensitivity of the ear was again determined for the rest of the audiogram. After this the incision was sutured and *S* was allowed to recover.

When a minimum of three months had elapsed, the exposed ear was again tested for sensitivity to tones from 100 cycles to 15,000 cycles. As a check on the general effects of the initial operation a 1-microvolt response audiogram was also determined at this time for the nonstimulated ear. No operative effects were observed. The control ear audiograms were similar to those of the exposed ear before stimulation.

When all measurements had been made, *S* was sacrificed and the temporal bones removed for later histological evaluation. The entire histological procedure has been described in detail previously (**2, 6**).

Five *S*s were tested and all were carried through histology.

Results

Electrical tests.—The immediate and more chronic effects of intense stimulation with a 10,000 cycle tone are

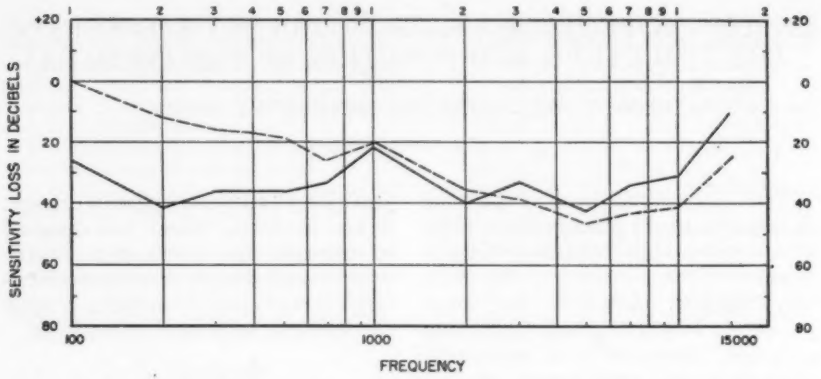


FIG. 1. Loss in sensitivity to an intense 10,000-cycle tone at two different points in time. The broken curve represents the mean loss measured immediately after stimulation. The solid curve presents this same group of ears more than three months later.

shown in Fig. 1. The immediate effects are widespread and show a progressive decrement in sensitivity from the lowest tone to the higher tones. Beyond the test tone of 10,000 cycles the sensitivity improves. This general picture coincides closely with that reported by Smith and Wever (7) in a study in which the measurements were made immediately after the overstimulation.

The more chronic effects show much the same pattern of loss for the high tones but a marked change in the response to the low tones. The curve has changed in the course of time so that now the effects of this stimulus are relatively uniform throughout the frequency range.

In Fig. 2 we have prepared the chronic individual loss curves for all *S*s. It can be seen that despite the fact that the same injury criterion of 40 db loss was used for each ear, these curves now vary widely in the magnitude of the loss of sensitivity. Two ears, 858 and 868, show a general improvement; one, 862, remains about

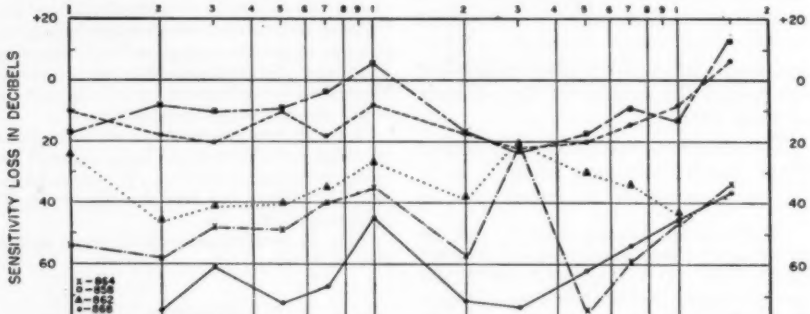


FIG. 2. Individual curves depicting sensitivity loss for all *S*s at least three months after injury. Sensitivity is measured here with each *S* serving as his own control.

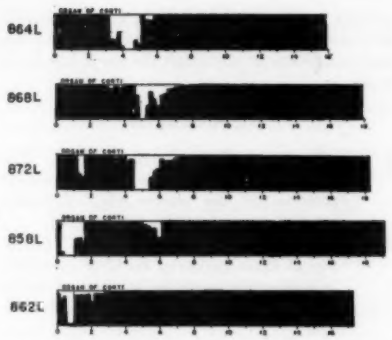


FIG. 3. Linear plots showing the extent and degree of injury to organ of Corti structures for each *S* in the study. Injury is represented by the unfilled space. The scale beneath each plot is laid out in millimeters from the base of the cochlea to the apex.

the same; while the remaining two, 864 and 872, continue to decline in sensitivity. What seems to be reflected here is that the power of recovery is an individual matter and is somewhat independent of the initial loss in sensitivity.

Histological tests.—The chronic effects of this intense high frequency sound on the structures of the organ of Corti are presented in Fig. 3. Here are the linear plots for all test animals. Included in the evaluation are estimates of the state of each of the hair cells and the supporting structures. In a particular judgment, 4/5 or 80% of the weight was given to the condition of the hair cells. The remaining weight was distributed among the other organ of Corti structures.

The lesion produced by a 10,000-cycle tone is in all cases confined to the basal turn of the cochlea. The primary center of damage is located between 4 and 6 mm. from the beginning of the basal turn, which is the same area found in the single case reported by Smith and Wever (7). In three of the cases a second area, between 1 and 2 mm from the base, also shows damaging effects. These results are puzzling but perhaps some light can be shed on them by examining the findings from two *S*s which were not originally part of this study. These *S*s were subjected to a tone of 2,000 cycles under the same procedure outlined earlier. When examined histologically, it was found that the injury

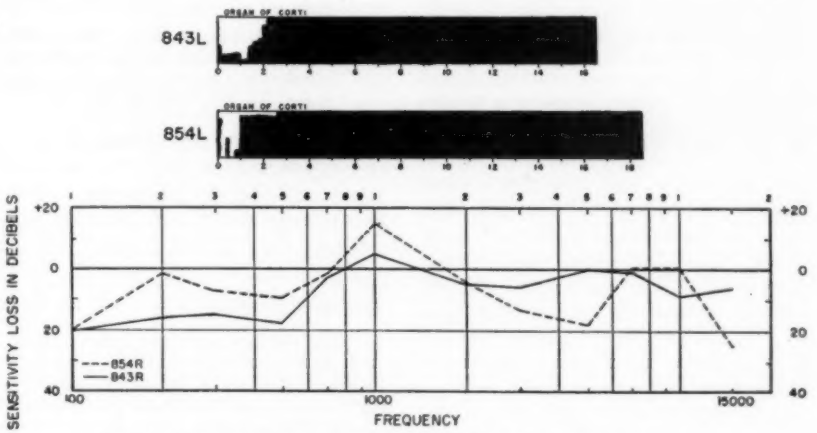


FIG. 4. Histological and electrophysiological results gathered from two ears exposed to an intense 2,000-cycle tone.

was centered in the very base of the cochlea, in the first 2 mm. Loss in cochlear potentials was not severe and did not seem to follow a regular pattern. These results are presented in Fig. 4. The suggestion here is that this area, the first 2 mm. in the base of the cochlea, is unusually fragile and will be affected by tones whose primary locus of damage is in the basal coil. If this is not the case, these instances must cause severe embarrassment for theories that hold to progressive tonal representation in the cochlea.

Summary

In this study we have shown that as time elapses between injury and test the effects of an intense high frequency stimulus are manifested almost uniformly on all tones in an audiogram reflecting electrophysiological sensitivity. This is in contrast to the initial effects of such a stimulus. At this time low tone sensitivity seems very little impaired.

The primary locus of injury, suggested by earlier investigators, has been verified by our results. In addition we have pointed out the possibility that structures in the very base of the cochlea may be extremely fragile and unusually susceptible to injury by tones whose primary damaging effect is in the basal turn.

The finding that a particular frequency, 10,000 cycles, injures a particular portion of the cochlea but does not affect the response to particular tones selectively, is in keeping with other stimulation deafness results. The implications are clear. Either one must contend that electrophysiological sensitivity as measured by cochlear potentials has very little to do with hearing, or else one must concede that the action of the cochlea is a general one in relation to stimulus frequency and could hardly account for the fine discriminations of pitch that the ear is known to make.

References

1. Alexander, I. E., & Githler, F. J. The effects of jet engine noise on the cochlear response of the guinea pig. *J. comp. physiol. Psychol.*, 1949, **42**, 517–525.
2. Alexander, I. E., & Githler, F. J. Histological examination of cochlear structure following exposure to jet engine noise. *J. comp. physiol. Psychol.*, 1951, **44**, 513–524.
3. Alexander, I. E., & Githler, F. J. Effects of intense pure tone stimuli when magnitude of initial injury is controlled. *J. exp. Psychol.*, 1953, **45**, 49–56.
4. Crowe, S. J., Guild, S. R., & Polvogt, L. M. Observations on the pathology of high tone deafness. *Bull. Johns Hopkins Hosp.*, 1934, **54**, 315–379.
5. Schuknecht, H. F., & Sutton, S. Hearing losses after experimental lesions in basal coil of cochlea. *Arch. Otolaryng.*, 1953, **57**, 129–142.
6. Smith, K. R. The problem of stimulation deafness. II. Histological changes in the cochlea as a function of tonal frequency. *J. exp. Psychol.*, 1947, **37**, 304–317.
7. Smith, K. R., & Wever, E. G. The problem of stimulation deafness. III. The functional and histological effects of a high-frequency stimulus. *J. exp. Psychol.*, 1949, **39**, 238–241.
8. Wever, E. G., & Smith, K. R. The problem of stimulation deafness. I. Cochlear impairment as a function of tonal frequency. *J. exp. Psychol.*, 1944, **34**, 239–245.

(Received July 15, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

EFFECTS OF TRAINING IN SPACE ORIENTATION ON PERCEPTION OF THE UPRIGHT[1]

MELVIN WEINER

Worcester State Hospital

The problem of this study was to determine how individuals under certain conditions could learn to reorganize their perceptions when they were given the opportunity to obtain new experiences. Specifically, when we set up experimentally a conflict between postural and visual cues, how do *S*s learn increasingly to utilize postural cues in the perception of the upright?

A systematic attempt to train people in this direction was carried out by Witkin (**5**). Here intellectual information was combined with demonstrations and practice in becoming better "informed" about bodily cues. While the training procedures did produce improved spatial orientation, Witkin concluded that it ". . . rarely altered the subject's perception of the situation. Its effect was rather to enable him to interpret his perceptual experiences in an intellectual way, and thereby arrive at a more nearly correct estimate of his own position and the position of the field" (**5**, p. 32). It was thus suggested that the ". . . training does not alter the impressions themselves, but relies on a more effective intellectual analysis of these impressions" (**5**, p. 33).

In this experiment we investigated whether learning to give greater weight to postural factors in the judgment of the upright could not only improve the judgment but also change the visual impressions accordingly. As in Witkin's study (**5**) *S*s were given information about the perceptual problems encountered in judging the upright when visual impressions and postural conditions were in conflict. However, in the present study, greater emphasis was placed on teaching individuals to orient themselves by *perceived body feelings* (e.g., perceived postural changes, pressures, and tensions), rather than by "figuring" and inferential analyses. New techniques directed toward helping *S*s to orient themselves in space *perceptually* rather than *intellectually* therefore were introduced. A reduction of the time interval from test to retest of about five weeks in Witkin's study to less than one week in the present study also was used to help decrease the possibility that the newly learned mode of orientation be forgotten. Finally, a luminous cube-like frame apparatus, viewed in a completely darkened room, was designed and used instead of the tilting room of Witkin's. The tilting room can be considered to be more articulately structured and stable than the cubic frame. The room is a visually more compelling framework. The luminous cube, on the other hand, is less visually compelling and less stable, and, therefore, a shift in view may be attained with greater ease.

It was hypothesized: (*a*) that *S* will learn to utilize postural experiences in relating his body to the objective upright, and that he will then more accurately estimate the objective verti-

[1] This study is based on a master's thesis presented to the Department of Psychology, Kansas University, and was carried out under the direction of Dr. Martin Scheerer. I am deeply grateful for his advice and suggestions in the research and for his help in preparing the manuscript.

cal in adjusting a luminous cube to the upright; (*b*) that the learning that accrues from the training will lead to real changes in the perceptions themselves in these orientation tasks; and (*c*) that this improvement will be generalized and will transfer to the task of adjusting a rod within the cube to the upright.

Method

Apparatus.—The apparatus consisted of a luminous coated rod surrounded by a luminous coated cube-like frame. This frame gave a three-dimensional appearance like the outline of a cube with the front part missing. The back was constructed of a 40-in. square frame of ¾ in. width; the sides of the "cube" were 47 × 1 in. strips of wood which projected from each corner of the frame toward *S*. The cube was mounted on a horizontal shaft which could be rotated to the left or to the right over a 360° range in the frontal-parallel plane. The rod, 39 × 1 in., was pivoted at the same center as the cube, but on a separate concentric shaft so that the rod and the cube could be rotated independently of one another. The pivot point of the rod and the cube was placed 8 ft. in front of *S* at eye level. An arrangement with protractors and stationary pointers permitted direct readings (in degrees) of the positions of the cube and the rod. When the rod was not in use, a black cover was placed over it. In this way either the cube alone, or the rod surrounded by the cube, could be presented to *S*. The cube and rod were coated with white luminous paint and during the experiment were the only items visible in the completely darkened room.

A tilting chair allowed *S* to be tilted in various positions. The completely padded chair was built with a high back, side supports, arm and foot rests, and an adjustable sponge rubber head rest, and could be tilted independently of the cube and the rod to a maximum of 38° to either side.

Procedure.—The following sequence of conditions was used: (*a*) *Pretest, Cube Test*: Positions of *S* and the initial position of the cube before adjustment of the cube were: *S*'s body tilted 28° L with the cube tilted L, R, L, R by 28°, and *S*'s body erect with cube again tilted L, R, L, R by 28°. (*b*) *Pretest, Rod-in-Cube Test*: *S*'s body was tilted 28° R, the rod and the cube also tilted 28°, but their positions varied as follows: (*i*) cube L, rod L; (*ii*) cube L, rod R; (*iii*) cube R, rod R; (*iv*) cube R, rod L. (*c*) One hour of training in space orientation (experimental group only). (*d*) *Retest, Cube Test.* (*e*) *Retest, Rod-in-Cube Test.*

The *S* was brought into the darkroom with his eyes closed and with masked aviator's goggles lowered over his eyes. He was seated in the upright chair and asked to lift his goggles and open his eyes on each trial. His task was to direct *E* how the cube (Cube Test) or the rod surrounded by the cube (Rod-in-Cube Test) should be adjusted from an initially tilted position so that they would look upright, i.e., straight up and down like the walls of the building or the flagpole outside. Between trials *S* was asked to lower a mask over his eyes and keep his eyes closed while the protractor was read. The test procedures followed that of the standard darkroom test as developed by Witkin and Asch (6).

At the end of the initial test series each *S* was interviewed and asked the following questions: "How did you determine when the cube (or rod) was upright? Did you have any special system? Did the cube (or rod) actually look upright, or was it just a matter of figuring it out?"

Several days after the two initial tests, *S* was taken to another room and was given 1 hr. of intensive training in space orientation. Each *S* was placed again in the experimental apparatus and at the appropriate times was given tape-recorded instruction in methods of taking better account of his body for judging the relationship of his body position to the objective upright. After *S* was seated in the chair the purpose of the training procedures was outlined to him. The *S* was informed that a critical attitude in looking at the cube was needed. He then was given training in judging his position in space, and was shown just what it felt like to be tilted at various angles when all visual impressions were eliminated, i.e., when his eyes were closed. Specific body cues were pointed out to help *S* become more aware of postural impressions when he was tilted, e.g., difference in pressure against the sides of his head from the head rest, pressure changes in his seat, pressure changes in his elbows, tension changes in his muscles, and feelings of total body tilt. The *S* then opened his eyes in the darkened room. While the cube remained in an upright position, the rod was placed in each of the positions in which *S* was tilted. The *S* thus was enabled to feel and at the same time see his position in space in relation to the objective upright. Following this training the rod was covered and no longer presented in any of *S*'s adjustments in the training session, and *S* was given the opportunity to make use of the newly experienced feelings of his position in space. The task was now that of adjusting the cube to the upright while *S* was in varying positions of tilt. After each judgment *S* received

information about its adequacy by showing him the cube in the objectively upright position. Thus, the procedure throughout was one of approximation and correction. In addition, *S* was instructed to visualize a vertical based on where he *felt* the vertical should be. He was given practice while tilted in different degrees in using the "felt vertical" as a basis for his judgments of the visual vertical. Throughout the training it was emphasized and demonstrated that his body gave him the "only" adequate basis for arriving at a correct visual orientation. As the following excerpt indicates, a body-conscious attitude, as opposed to an intellectual attitude, was stressed. "People sometimes misunderstand what we mean by figuring out the position of the cube and rod on the basis of their bodily sensations. They think we mean a kind of intellectual computation of angles and degrees of their body tilt. That is not what we mean. What we mean is that you must actually become more sensitive to your body sensations. The reason we say to go by body sensations is that your body is an excellent judge of where gravity is. Therefore, you must really feel the muscle tensions in your body. In these situations you must, so to say, think with your body and not with your brains alone. Just thinking and figuring with your brains is not enough. In fact, taking just an intellectual attitude and going by blind figuring may make you a poorer score. If you truly go by the feelings in your body, your score will improve."

Several days after the training session *S*s were retested on the two identical initial tests (Cube and Rod-in-Cube Tests).

In order to control the factor of initial test experience, as contrasted to specific training experience, a control group was tested on the same initial tests and retests, but without the intervening training. The only difference was that the control group was not interviewed at the end of the *initial* tests, so that these *S*s would receive no further cues.

In order to decrease forgetting effects, an attempt was made to obtain the shortest possible time interval between the different testing conditions. There was an average of five days between pretest and training session, three days between training session and retest for the experimental group, and two days between the pretest and retest for the control group.

There were 25 *S*s (13 male and 12 female) in the experimental group and 20 *S*s (10 male and 10 female) in the control group. All *S*s were University of Kansas students who volunteered for the experiment. If *S* missed the appointment for the training session or spontaneously verbalized dislike for the experiment on the pretest, he was dropped and a replacement obtained. If *S* appeared for the training session but did not come for the retest, *E* sought him out and asked him to complete the retest. In all, three *S*s were dropped for missing appointments, and one *S* for dislike.

Results

The mean errors of the experimental and control groups on the pretests are presented in Table 1. The amounts by which each *S*'s adjustments of the cube or rod deviated from the true upright were averaged without regard for sign, and means and *SD*'s were obtained for all *S*s in each group. All differences between the experimental and control groups were nonsignificant. Thus, the control group may be considered adequate.

Effects of training on Cube Test performance.—The effects of training are measured by the improvement from pretest to retest in mean errors made by the experimental group corrected for the improvement (if any) made by the control group. These values are presented in Table 2, and indicate that in adjusting the cube with body tilted the experimental group improved significantly more than the control group. The variances of the difference scores between the two groups were significantly different, as indicated by the *F* test. Therefore,

TABLE 1

Mean Errors in Pretests

Condition	Experimental (N = 25)		Control (N = 20)		t
	Mean	*SD*	Mean	*SD*	
Cube Test					
Body 28° L	13.1°	7.8	11.9°	4.2	0.60
Body erect	4.9°	3.4	4.2°	2.8	0.75
Rod-in-Cube Test					
Body 28° R	15.7°	10.5	15.0°	7.6	0.25

TABLE 2

MEAN IMPROVEMENTS IN RETESTS

Condition	Experimental (N = 25)		Control (N = 20)		t
	Mean	*SD*	Mean	*SD*	
Cube Test					
Body 28° L	6.7°	5.4	3.2°	3.0	2.65
Body erect	1.7°	4.8	0.4°	2.3	1.14
Rod-in-Cube Test					
Body 28° R	5.1°	7.6	0.2°	2.7	3.00

the approximation method, as described by Edwards (1), was used. The mean improvement difference of 3.5° between the two groups is significant between the 1% and 2% levels. This reflects a decrease of the influence of the visual field and greater reliance on bodily cues after training. In terms of mean errors for each trial, the experimental group improved significantly by 6.7° in dropping from a mean error of 13.1° to 6.4° (51%). The control group also improved significantly by 3.2° in dropping from 11.9° to 8.7° (26%).

When adjustments were made from an *erect* position, the experimental group did not improve much more than the control group. The mean improvement difference between the two groups is not significant. Furthermore, even when the improvement in each group is considered separately, neither group improved their performance significantly in an upright position.

Effects of training on Rod-in-Cube Test performance.—The effect of training in bodily orientation which the experimental group received in the Cube Test on performance in the Rod-in-Cube Test is also shown in Table 2. The mean errors of the experimental group improved by 5.1° in dropping from a mean error of 15.7° to 10.6° (32%), while the control group improved by only .2° in dropping from 15.0° to 14.8° (1%). The mean improvement difference of 4.9° between the two groups is significant ($p < .01$). This suggests greater use of postural experiences and greater ability to judge the rod independently of the surrounding visual framework after training. Further, while the control group by itself showed no significant improvement on the retest, the experimental group by itself showed improvement which is significant ($p < .01$). Thus, the training procedures in the cube situation seem to have led to the acquisition of a general principle which was applied to the different orientation problem of the Rod-in-Cube Test. The previous initial test experience obtained by the control group was limited in effectiveness to the cube retest.

Individual differences.—While the results have indicated that training to utilize bodily experiences leads to improvement in orientation performance, there were important individual differences in the degree to which this was true. For example, in the Rod-in-Cube Test, at one extreme, one *S*'s performance became even more visually dependent by 6.0° after training; at the other extreme, another *S*'s performance improved by 28.5° after training.

Qualitative analyses.—The following comments illustrate the methods used by *S*s on the *retest* after training: "A *feeling* of being at various degrees (of tilt) was most important." Another *S* stated, "Originally I had tried to use this kind of system naturally, but I did not have such a good idea as to my degree of tilt." Another *S* explained, "I would *feel* at which angle my body was. Then I tried to tell where upright should be in relation to my body. I also used a visual vertical in my mind between trials. With the rod I used the same method, but regard-

less of the cube. *Things looked differently this time.* When I made my adjustments, they (the cube and the rod) looked and felt straight."

These reports suggest that in order to make effective use of bodily experiences in judging the position of the cube or the rod, *perceiving the position of one's own body in space* is a primary factor.

Increased phenomenal stability.—Further evidence of the effects of training on *S*'s initial mode of orientation was the tendency for the perceived stability of the visual framework to increase after training. Perceived stability of the visual framework is indicated by a decreased frequency in the occurrence of perceptual shifts of the main axes of the cube. Perceptual instability of the cube axes reflects *S*'s insufficient reference to bodily experiences in creating an adequate frame of reference in determining the upright.

After training there was a decrease from two to one perceptual shifts on the Cube Test, and when it did occur, the mean error was reduced from 40° to 16°. After training there was a decrease from four to zero perceptual shifts on the Rod-in-Cube Test. For example, for one *S* the mean errors were reduced from 66° to 4° and from 63° to 12°. Moreover, on both tests in no case did a perceptual shift occur in a retest where it did not occur in the pretest. These results suggest that when an individual can more precisely perceive the position of his body in space and make greater use of his bodily experiences, the visual framework becomes more stable.

Perception vs. "intellectual analysis."—To ascertain whether the noted improvements in orientation represent changes in the perceptions themselves, or whether training improved performance indirectly by the use of "intellectual" interpretations, immediately after the retest *S* was asked, "Did the cube (or rod) actually appear upright, or was it a matter of figuring it out? In other words, if you had wanted to make the cube (or rod) actually look upright, would you have left it in the position to which you adjusted it, or would you have changed its position still further?" The responses derived from these instructions are called first retest responses.

If *S* reported that he was not sure whether his final adjustments actually appeared upright to him, or if he thought that he had used some kind of "figuring," the following special instructions were given: "We will now repeat the same procedure, but this time say 'stop' when the cube (or rod) actually appears upright to you." After each response, *E* asked, "Does the cube (or rod) now look upright? Does it look the same as what you think to be upright, or different?" (Second retest responses.)

The results with the cube were that 19 *S*s (76%) at first reported that their judgments coincided with what actually looked upright; an additional 5 *S*s (20%) reported in the second retest responses that their judgments coincided with what actually looked upright. The total for actual perceptual responses with the cube is, therefore, 24 *S*s (96%). With the rod the corresponding results are 13 *S*s (52%) for first retest responses and 6 *S*s (24%) for second retest responses, making a total of 19 *S*s (76%) where the judgments of the rod coincided with what actually looked upright. With 5 *S*s (20%) the judgments with the rod were inconsistent, and sometimes coincided and sometimes did not coincide with the way the rod actually looked. In only one case in the Cube and Rod-in-Cube Test did improvement result from intellectual or inferential analysis of the situation and not from a change in the perceptual impressions themselves. It thus is indicated that most *S*s in both tests judged the vertical more correctly after training and also arrived at an actual perceptual adjustment.

The effects of increasing experience with the environment on perceptual change and development were studied in more detail in the case of one male *S* who volunteered for further experiments. In a second training session *S* was given further practice in estimating body position with eyes closed and several trials in estimating the position of the cube. He was retested after this training and

also was given an additional retest eight months later. Not only were there large improvements after training, but the final mean errors were very small. The results on the pretest and three retests were 39.8°, 13.3°, 3.8°, and 6.0° with the Cube Test, body tilted; 9.5°, 2.5°, 3.8°, and 5.8° with Cube Test, body upright; 20.8°, 7.5°, 3.3°, and 3.5° with Rod-in-Cube Test, body tilted. These scores reflect effective use of bodily experience and judgments made independently of the visual impressions. The *S* reported that the cube and the rod actually *appeared* upright on all tests. The magnitude of these changes in mode of orientation is very striking and demonstrates the effects that learning to rely on postural experiences may have.

Discussion

The results indicated that there was no significant improvement in orientation when the body was in an upright position. It should be noted, however, that the process of adjusting the visual field (the cube) to the upright when the body is erect is relatively simple, since the upright position of the body makes for the easiest determination of the true upright. All the task requires is lining up the cube with the upright body. No *S*s had difficulty in feeling that they were sitting erect. The various training procedures directed toward helping *S* to determine his position in space would, therefore, be expected to have a lesser influence on the judgments made in the upright position since this position already provided optimal body cues for *S*.

The training procedures appeared to have helped *S*s to experience their body postures in a new way, a new mode in which body feelings were emphasized. Consequently, *S*s were better able to determine their own position in space. A more accurate determination of an individual's position in space does not *only* involve a "knowing" of a certain position in relation to the vertical, but also involves a newly learned "feeling" of where the gravitational upright is in relation to the experienced feelings in the individual's body.

Evidence that the feelings within the person's body are an important factor in the determination of the objective vertical is also presented by Mann, Berthelot-Berry, and Dauterive (**3**). In their experiment precision of judgment was increased when nonlabyrinthine proprioceptive cues were heightened by the elimination of the padding in the sides, seat, and back of the tilting chair.

It appears that new experiences may play an important role in the development of a new mode of perception. The significantly greater improvement of the experimental group on the Cube Test and its transfer to the Rod-in-Cube Test seem to have been accomplished by learning actively to utilize bodily cues in a more sensitive way in relating the body to the objective upright. In other words, *S learned to perceptually reorganize the potency given postural and visual cues*. These results may be compared with those of other studies (**2, 4**), where evidence of reorganizational perceptual learning also was presented.

In the present study we have focused on changes in the perceptual achievements of *S*s. Since the same perceptual outcome can be achieved through different underlying perceptual processes, further study focused on the *intervening or mediating processes*, such as the body image or schema, which would more completely account for changes in perceptual outcome, is needed. Furthermore, the training in the present study was "laboratory-bound." Experiences which are less directive and less laboratory-induced, and which would require more active participation from *S*, would help to increase the representativeness of the situation and thus help to increase the generality of the results. In addition, we found striking individual differences in the degree to which individuals could learn to restructure their preferential modes of orientation in perceiving the upright. These findings indicate the need to explore the dimension of flexibility-rigidity of the perceptual system. In other words, to what extent can different persons make use of new experiences in changing their preferen-

tial mode of perceiving—which includes the problem of perceptual unlearning as well as perceptual learning?

Summary

The present study attempted to teach *S*s how to effectively utilize postural experiences in the perception of the upright. An experimental group (25 *S*s) and control group (20 *S*s) were given the following pretests: (*a*) With *S* tilted, a luminous cube in a completely darkened room was adjusted to the true upright; (*b*) The same task for *S* in an upright position; (*c*) With *S* tilted, a luminous rod surrounded by this cube was adjusted to the true upright. Subsequently the experimental group alone was given 1 hr. of specialized training in space orientation, employing various positions of body tilt and emphasizing bodily cues. Both groups then were retested on the same original tests.

No significant improvement occurred in either group with the body upright in the Cube Test, though training led to significant improvement in the Cube Test with the body tilted. While there also was significant improvement in the control group upon retesting in the Cube Test, the improvement shown by the experimental group was significantly greater than that of the control group. The training also led to significant improvement in the experimental group in the Rod-in-Cube Test, which served as a test of transfer. In contrast, no significant improvement was shown by the control group in the Rod-in-Cube Test. The noted improvements were attended by increased stability of the visual field, and in most cases, represented real changes in the perceptions themselves. Marked individual differences in the degree to which training led to improvement in orientation also were noted.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Kilpatrick, F. P. Two processes in perceptual learning. *J. exp. Psychol.*, 1954, **47**, 362–370.
3. Mann, C. W., Berthelot-Berry, N. H., & Dauterive, H. J., Jr. The perception of the vertical: I. Visual and non-labyrinthine cues. *J. exp. Psychol.*, 1949, **39**, 538–547.
4. Weiner, M. The development of perception in adults. Topeka, Kans.: Menninger Foundation, 1954. (Mimeo.)
5. Witkin, H. A. The effect of training and of structural aids on performance in three tests of space orientation. Washington, D. C.: Civil Aeronautics Administration, 1948. (Rep. No. 80.)
6. Witkin, H. A., & Asch, S. E. Studies in space orientation. IV. Further experiments of perception of the upright with displaced visual fields. *J. exp. Psychol.*, 1948, **38**, 762–782.

(Received July 28, 1954)

Journal of Experimental Psychology
Vol. 49, No. 5, 1955

TASTE EFFECTS RESULTING FROM INTERMITTENT ELECTRICAL STIMULATION OF THE TONGUE [1]

ROSEMARY PIERREL

Brown University

Intermittent stimulation in other sensory fields has been employed to assess the time constants of the receptor process. The critical fusion function, in the case of vision, describes those frequencies of intermittence and intensities of stimulation at which the sensory report changes from steady to interrupted, or vice versa, depending upon the direction of stimulus change. The form of such critical fusion functions in vision reveals a duplex character which is attributed to the duplicity of the sense organs, i.e., the rods and cones.

Since the normally adequate taste stimuli consist of solutions, it is almost impossible to control sharply the period of their stimulating effect. The use of inadequate electrical stimuli has been suggested as a means of effecting this control and, in particular, to permit intermittent stimulation. Allen and Weinberg (1) utilized intermittent electrical stimuli for this purpose. They attempted to determine the critical fusion frequency for the sour taste. The cathode, the stimulating electrode, was placed on the dorsal surface of the tongue. Various intensities of stimulating square wave currents were employed. The intensity was fixed and the intermittence rate adjusted up to the point at which *S* reported he could no longer discriminate an intermittence, i.e., stimulation appeared to *S* to be continuous. When several hundred individual measurements for different voltages were plotted, frequency against voltage, the points were found to fall along four well-defined curves. These findings indicated that as the potential increases the CFF is reduced. The fact that there were four curves was interpreted as strong evidence for the existence of four fundamental taste mechanisms.

More recently Jones and Jones (4), attempted to reproduce this portion of the Allen and Weinberg experiment. Their effort to obtain fusion values upon intermittent cathodal stimulation was unsuccessful. In no case were their *S*s able to detect any discontinuity of the stimulus. Thus, it was impossible to obtain critical fusion frequencies of taste. Jones and Jones speculated that Allen and Weinberg's *S*s may have been responding to a fusion of cutaneous sensations.

At its inception the present experiment was designed to parallel that of Allen and Weinberg. In their study, frequency was varied while intensity was held constant. However, in our situation it was more feasible to hold frequency, or intermittence rate, constant and to vary intensity. Another change in the stimulating procedure was to use the anode as the stimulating electrode. Cathodal stimulation elicited pain which obscured other effects.

As the intensity of an intermittent anodal stimulus on the tongue is increased, *S* first detects stimulation as

[1] Submitted in partial fulfillment of the requirements for the Ph.D. degree in the Department of Psychology at Brown University. This research was supported in part under Contract N7 onr-35804 between Brown University and the Office of Naval Research. The writer wishes to acknowledge her indebtedness to Professor Carl Pfaffmann for his help and guidance throughout the course of this work.

a sour or metallic steady taste. The threshold for this response is termed the electric *taste* limen. With a further increase in intensity a point is reached at which intermittence or pulsation appears. This is called the critical fusion or *pulse* threshold. In the first part of this study the relation of taste threshold to pulse threshold was determined at the same point on the tongue. In the second part of this study the effect of the same stimulus on the lip, a taste insensitive area, was studied.

Method

Trains of square wave stimuli of varying frequency were obtained by leading the output of a Hewlett-Packard, Type B audio oscillator through a limiting circuit (6). This consisted of an overloaded triode to produce peak clipping, and a cathode follower in which the cathode resistor served as a potentiometer to provide different voltages for *S*. A reversing switch in the output to *S* made it possible to reverse the polarity of the tongue electrode with respect to the indifferent electrode.

Possible variations in voltage output in the apparatus were monitored with an AC voltmeter, calibrated in arbitrary units for the square wave signal. The entire circuit was calibrated with a cathode ray oscillograph. Intensity units in the following graphs are stated in potentiometer dial units in which 10 units were equal to .57 peak volts.

The *S* was seated in a straight chair with the head supported by a forehead and chin rest. A mirror was placed in front of *S* at "tongue level" which enabled him to select the location for stimulation and to arrange the contact of the electrodes. The cathode consisted of a circular piece of pure silver about 2.5 cm. in diameter by 1 mm. in thickness. The cathode was placed inside *S*'s left cheek. The anode consisted of a silver ball approximately 1.5 mm. in diameter, rigidly supported in a holder so that *S* could place his tongue or lip against the electrode.

Trials consisted of a series of training sessions and two different experimental series in which limens for electrical stimulation were obtained. The same three *S*s were employed in both experiments. This group consisted of two male graduate students (A. M., A. R. T.) and one female undergraduate (J. S.), ranging in age from 16 to 25 years.

Eight frequencies of stimulation were applied to the tongue using the method of limits. The intensity of stimulation was increased from zero up to the value at which *S* indicated the sensation to be just supraliminal (ascending series), or the stimulus intensity as initially presented was at a point well above the limen and progressively decreased until *S* signalled "sensation gone" (descending series). Each session consisted of a total of 64 trials. The trials were presented in blocks of four at a given frequency in an ascending-descending-descending-ascending order. The range of frequencies employed was from 20 to 1000 cycles/sec. Only one-half of the total number of frequencies was presented on a given day, the other half being presented on the following day. The order of frequency presentation was as follows: 1000, 20, 500, 30, 30, 500, 20, 1000; or 300, 50, 200, 100, 100, 200, 50, 300.

The electrodes were polished with silver polish and wiped with 70% alcohol both before and in the middle of each session. The *S* was asked to draw the tongue back into the mouth between successive trials. Presumably the washing action of the saliva would eliminate polarization products which might accrue during stimulation. Although the intertrial interval was not strictly timed, the period between the beginning of one trial and the beginning of the next, effectively, was slightly less than 60 sec.

Instruction.—The *S*s were informed that this was an experiment on the sense of taste and that we were interested in finding out how a slight electric shock "felt" under different conditions. They were requested to select a clearly visible papilla on the anterior dorsal surface of the tongue and to take care that the electrode was accurately placed, centered on this particular papilla, not only throughout a session but also throughout a particular experimental series, extending over a period of 6 to 10 days. Tongue maps were drawn to aid in this localization process. The A-D-D-A order was explained and *S*s agreed to signal in response to *E*'s "ready" when correct electrode placement had been made. This was followed by *E*'s announcement of "ascending" or "descending." The *S* again signalled when the stimulus "came in" or "went out."

Experiment I.—Six days of combined taste and pulse trials were carried out in which the same papilla was employed for both types of judgment. One-half of all frequencies employed was presented on a given day, the other half the next day, and so on in straight alternation. Taste and pulse limens were obtained successively at each frequency; e.g., four taste trials at 1,000 cycles were immediately followed by four pulse trials at 1,000 cycles, then four taste trials at 20

cycles were followed by four pulse trials at 20 cycles, and so on through the frequency series.

Experiment II.—In this experiment the lip limens were obtained by placing the anode on the mucous inside surface near the middle of the lip. Although *S* attempted to place the electrode in the same spot each time it is probable that placement was less accurate than on the tongue owing to the lack of papillae to serve as landmarks. However, it is not likely that this introduced undue variability. A limited number of observations was carried out in which pulse limens were obtained on each *S* with the anode placed at several different locations on the lip mucosa. These limens did not appear to vary within *S*s as a result of different localizations of stimulation.

This experiment consisted of 10 days of experimental sessions. In view of the not inconsiderable changes in posture required for *S* to shift from the tongue to lip electrode position the tongue trials for all frequencies for a given session were carried out followed by the lip trials. At the next session the lip trials preceded those on the tongue. One-half of all frequencies employed was presented on a given day. The other half was presented on the succeeding day.

Sessions comparing reversed polarity of the electrodes were attempted. In this situation taste and pulse limens were obtained from a given papilla as in Exp. I. Next a reversing switch in the clipping circuit was thrown which caused the tongue electrode to become the cathode. A systematic investigation comparing these limens was made impossible as *S*s reported "pain" at the site of the cathodal tongue electrode. Therefore, these sessions were discontinued after only a few trials.

Results

Experiment I.—Anodal stimulation at close to liminal intensities (taste threshold) was described predominantly as sour and metallic. Sour was most characteristic of the lower frequencies, but in all instances the taste was described as steady and continuous. Anodal stimulation at higher intensities produced a heightening of the steady taste up to the point where an intermittent pulsation appeared (pulse threshold). A burning quality was sometimes described at this level.

The taste and pulse threshold

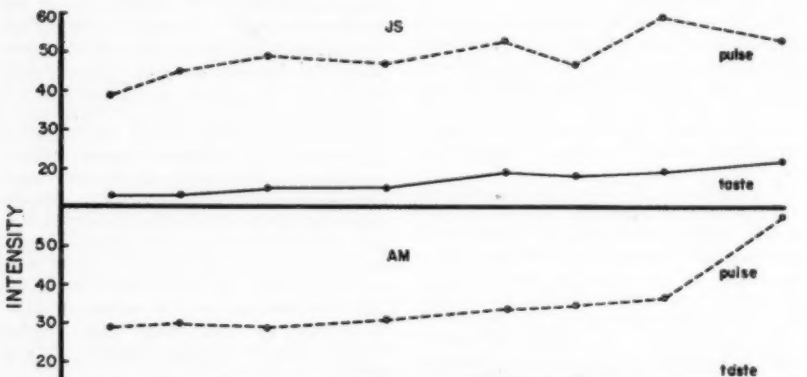


FIG. 1. Comparison of taste and pulse limens: Limen intensity as a function of frequency.

curves for the three *S*s are shown in Fig. 1. Each point on each curve represents the mean value of 24 trials. In the case of both taste and pulse there appears to be some tendency for the limen to increase with increase in frequency. The electric taste threshold values are below the electric pulse threshold values for any particular point on the tongue. By means of a simple visual comparison it is difficult to determine whether or not a given *S*'s two curves are similar in form. Therefore, a test for the significance of trend differences was undertaken. The trend analysis is an extension of analysis of variance techniques outlined by Lindquist (**5**). In the present instance it was applied to answer two questions: (*a*) In the case of any given *S*, are the two curves similar or identical in form? (*b*) If they are similar in form do they lie in essentially the same place on the ordinate?

The hypothesis that for each frequency taste and pulse limens all have the same population means is tested by

$$F = \frac{\text{Limen type} \times \text{Frequency}}{\text{Location} \times \text{Frequency} \times \text{Day of testing}}.$$

If F is significant the hypothesis as stated may be retained, that is, a significant F ratio in this situation indicates that the two curves being compared are significantly different in form. If the results of the first F test indicate that the curves are essentially similar in form (F is nonsignificant), then a second test may be made. This is a test of the hypothesis that the mean of the differences in taste and pulse means is zero for the entire population, or that the two curves are essentially coincident on the set of coördinates. On the assumption that there is no real Limen type × Frequency interaction the second hypothesis may be tested by

$$F = \frac{\text{Limen type}}{\text{Limen type} \times \text{Day of testing}}.$$

If this ratio may be shown to be significant we may conclude that the curves are not coincident on the ordinate.

TABLE 1

TREND ANALYSIS: TASTE AND PULSE CURVES

S	*F* Ratios					
	$\frac{L \times F}{L \times F \times D}$	*df*	*p*	$\frac{L}{L \times D}$	*df*	*p*
A. M.	$\frac{.01260}{.00180}$	$\frac{7}{14}$	.01			
A. R. T.	$\frac{.00290}{.00090}$	$\frac{7}{14}$	.05			
J. S.	$\frac{.00428}{.05961}$	$\frac{7}{14}$	—	$\frac{2.71960}{0.01770}$	$\frac{1}{2}$	.01

The results show that for two *S*s, A. M. and A. R. T., the F ratio for trend is significant. This indicates that for each of these two *S*s the taste and pulse curves are not parallel. Trend analysis of the data for J. S. indicates that her curves lie parallel to each other, although they are not coincident on the same set of coördinates.

Experiment II.—Figure 2 shows the results obtained when intermittence limens were secured from the tongue and the taste-free lip mucosa during the same experimental session. Each point on each function represents a mean of 40 observations.

The trend test for the hypothesis that pulse limens for tongue and lip are identical at each frequency is

$$F = \frac{\text{Location} \times \text{Frequency}}{\text{Location} \times \text{Frequency} \times \text{Day of testing}}.$$

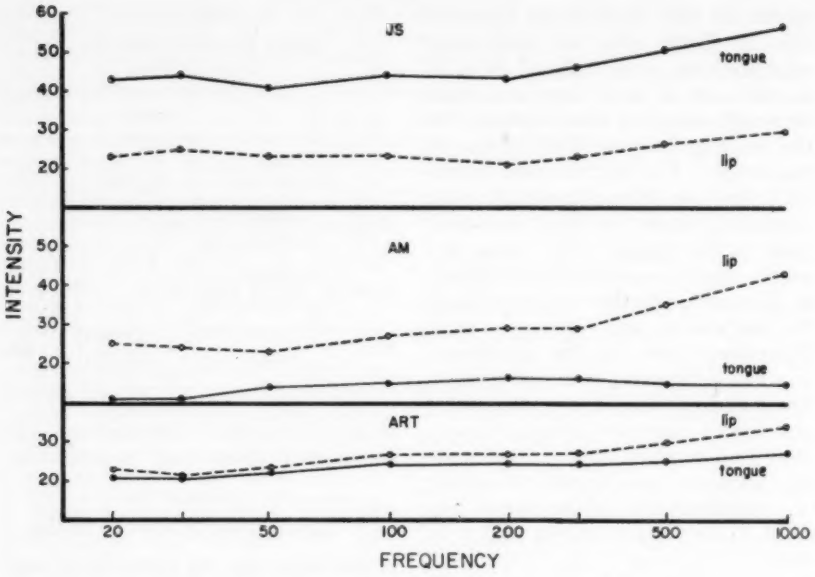


FIG. 2. Comparison of pulse limens on tongue and lip: limen intensity as a function of frequency.

and the test whether the two curves plotted on a set of coördinates coincide is

$$F = \frac{\text{Location}}{\text{Location} \times \text{Day of testing}}.$$

The results of the trend analysis indicate that for all three *S*s the tongue and lip functions are not significantly different from each other in form. With the exception of A. R. T., the *S*s' two curves are shown to lie in disparate regions along the ordinate. This difference in location for the tongue and lip functions is shown to be significant at better than the 1% level of confidence. As one might expect from visual inspection of the mean functions, the curves for A. R. T. are not only similar in form but also essentially coincident.

An analysis of the tendency for the 1,000-cps limen to exceed that of the 20-cps limen in the pulse conditions was made. When both experiments are taken together there are three mean pulse threshold values at each frequency for each of the *S*s, i.e., one set from Exp. I and two from Exp. II. This gives a total of nine sets of pulse limen means. In each of these nine cases the limens at 20 cps and at

TABLE 2

TREND ANALYSIS: TONGUE AND LIP CURVES

S	*F* Ratios					
	$\frac{L \times F}{L \times F \times D}$	*df*	*p*	$\frac{L}{L \times D}$	*df*	*p*
A. M.	$\frac{.0773}{.2066}$	$\frac{7}{28}$	—	$\frac{2.0867}{0.1573}$	$\frac{1}{4}$	.01
A. R. T.	$\frac{11.54}{25.16}$	$\frac{7}{28}$	—	$\frac{189.12}{51.05}$	$\frac{1}{4}$	—
J. S.	$\frac{.00195}{.00475}$	$\frac{7}{28}$	—	$\frac{1.4891}{0.0127}$	$\frac{1}{4}$	.01

Note.—With a single exception all *S*s' data for both experiments were transformed into logarithms in order to meet the requirement of homogeneity for analysis of variance. The data for A. R. T. in Exp. II did not require such transformation.

1,000 cps were examined to determine whether the limen value at 1,000 cps exceeded that of the limen at 20 cps. It was found that in every case the 1,000-cps limen exceeded that at 20 cps. Tukey's signed rank test for paired replicates (7) was applied to these data. Since in all nine of our cases the 1,000-cps limen is higher we may state in terms of this test at better than the 1% level of confidence that the 1,000-cps limen is higher than the 20-cps limen in a significant number of cases.

Discussion

The finding that the relation between frequency and intensity for the anodal taste threshold is essentially a continuously rising curve is partial confirmation of an earlier experiment of Bujas and Chweitzer (3) which employed a range of lower frequencies. For the frequency range common to both experiments (30–1,000 cps) these investigators also obtained a continuously rising curve. Bujas (2) also determined taste fusion thresholds employing anodal square wave currents between .5 and 10 cps. He indicated that upon increase in intensity at any one frequency the report "continuous sour" became one of "intermittent sour." The taste fusion frequency was found to increase with an increase in intensity. The flickering tastes obtained by Bujas appear to be different in nature from the impressions obtained at pulse threshold in the present investigation. These latter appeared to be superimposed upon a background of steady taste perception. In other words, the taste itself did not pulsate, rather a new intermittent component was added to the existing taste. The disparity in the nature of this pulsation and the intermittent taste of Bujas may be attributable to the differential effects of high and low frequencies.

Jones and Jones, employing cathodal stimulation, and the writer, stimulating with the anode, were never able to obtain a report of an intermittent flickering taste in the frequency range 20–1,000 cps. In the present experiment the pulse threshold was obtained at intensities substantially above the taste threshold. The reports are, as noted, characteristically of pulsation or vibration. The reports indicate that no impressions of sourness, metallic or other taste perception occurred when the lip was stimulated. On the lip the only report made was the vibratory one. Thus, no taste threshold could be obtained. The *S* reported pulsation as the only effect of stimulation. It was possible, however, to obtain pulse threshold curves from a taste-insensitive area which were essentially similar to those obtained in a taste-sensitive area. The finding that the pulse threshold-frequency relation is similar on tongue and lip suggests that this is a vibratory-tactile threshold rather than an aspect of taste.

The question now arises as to whether Allen and Weinberg were also producing vibratory rather than taste effects by cathodal stimulation. In their experiment the tongue was stimulated with intermittent electrical stimuli. Frequency was employed as the dependent variable and fusion frequencies were obtained for a range of set intensities. They found, as has been noted, that the critical fusion frequency decreased with increase in intensity. In the present experiment, however, thresholds increased with frequency. In this respect the findings of the two experiments are at variance.

Other investigators (1, 3) report rather great variability in the thresholds to intermittent electrical stimulation of the tongue, whether for taste or for pulsation threshold. This has been noted in the present experiments but this variability was not great enough to mask certain trends and relationships.[2] Allen and

[2] Geldard, in a personal communication, has pointed out that since the effective stimulus is probably current flow, in this case the specification of stimulus magnitude by voltage is open to the criticism that variation in *S* and electrode resistance would lead to change in apparent stimulus voltage, whereas the threshold specified in current values might show little variation. This point was checked during a series of experiments on strength-duration relationships in electric taste stimulation after the present paper was prepared for publication. Taste thresholds were obtained both with and without the inclu-

Weinberg noted this variability and stated that not until all the data were plotted could any systematic relationships be observed. They then found that all the points fell on one of four discrete curves.

Some of the anodal pulse data obtained in the present investigation were treated in the same way. That is, the individual values obtained at each session were plotted in the form of a frequency distribution at each intensity and frequency. We could observe no tendency for the points to fall along discrete curves or regions of the graph.

As the intensity of an intermittent stimulus on the tongue is increased the sensory report changes from steady to pulsating. However, with further increases in intensity there is no reversion to steady. The pulse thresholds here reported represent the only pulse threshold which could be obtained. Therefore it is suggested that these pulse thresholds may be equivalent to the critical fusion frequency thresholds obtained by other investigators from the tongue. The pulse threshold curves obtained in the present experiment are believed to be intensity-frequency curves for vibratory-tactile sensitivity. This is based on the nature of the subjective reports and upon evidence from the tongue and lip comparisons. If the Allen and Weinberg determinations are of the same nature the fact that intensity decreases with frequency might be explained by the different conditions of experimentation.

Summary

The use of intermittent electrical stimulation on the tongue makes it possible to examine relationships between frequency and intensity of stimulation and pulse or "critical fusion frequency" of taste. Experiments were conducted in which the tongue was stimulated by square wave currents at frequencies between 20 and 1,000 cps. Two types of limen to anodal stimulation were found employing the method of limits. Thresholds were obtained from the tongue and lip.

From the results of the present experiment the following conclusions may be drawn:

1. Two types of threshold to a square wave intermittent electrical stimulus can be obtained from the tongue: an electric taste threshold and an intermittence or pulse threshold. The electric *taste* described by the *Ss* did not itself appear to fluctuate. Such intermittence or vibration as was discriminable appeared to be either superimposed upon, or substituted for the steady taste; *Ss* sometimes characterized the latter in such terms as "pure vibration."
2. Electric taste threshold values are below the electric pulse threshold values for any particular point on the tongue.
3. No electric taste thresholds could be obtained from the lip as there was no discriminable taste. Intermittence thresholds were obtained.
4. Both electric taste and electric pulse thresholds tend to rise as frequency is increased.
5. Similar electric pulse threshold-intensity-frequency functions may be obtained on the tongue and on the lip.
6. It is believed that pulse threshold functions are equivalent to critical fusion frequency curves of taste. If this is the case they are essentially vibratory or cutaneous pulse threshold functions. Therefore, inferences regarding receptor mechanisms of *taste* cannot be drawn from pulse threshold functions.

References

1. Allen, F. & Weinberg, M. The gustatory sensory reflex. *Quart. J. exp. Physiol.*, 1925, **15**, 377–383.
2. Bujas, Z. Quelques données sur le goût electrique. *Année Psychol.*, 1949, **50**, 159–168.
3. Bujas, Z., & Chweitzer, A. "Goût Electrique" par courants alternatifs chez l'homme. *C.R. Soc. Biol.*, Paris, 1937, **126**, 1106–1109.
4. Jones, M. H., & Jones, F. M. The critical frequency of taste. *Science*, 1952, **115**, 355–356.
5. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
6. Seely, S. *Electron-tube circuits.* New York: McGraw-Hill, 1950.
7. Wilcoxon, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid, 1949.

(Received July 29, 1954)

sion of a 560K ohm resistor placed in the circuit in series with *S*. This increased resistance rendered the resistance changes in *S* less effective. However, the relative variability in the two conditions was approximately the same. These findings suggest that the major factors producing variability of threshold did not arise from changes in *S* or electrode resistance.